Process modeling for proximity effect correction in electron beam lithography

Thiago Rosa Figueiro

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Nano Electronique Nano Technologies**

Arrêté ministériel : 7 août 2006

Présentée par

Thiago ROSA FIGUEIRO

Thèse dirigée par **Jean-Herve TORTAI** et **Patrick SCHIAVONE**

préparée au sein du **Laboratoire des technologies de la microélectronique** dans **l'École Doctorale Electronique, Electrotechnique, Automatique et Traitement du Signal**

Modélisation des procédés pour la correction des effets de proximité en Lithographie électronique

Thèse présentée et soutenue à Grenoble le 19/01/2015 devant le jury composé de:

Christophe VALLEE – Président du jury
Dominique DROUIN – Membre du Jury
Régis OROBTCHOUK – Membre du Jury
Bertrand IOOSS – Membre du Jury

“My most brilliant achievement was my ability to be able to persuade my wife to marry me.”

- *Winston S. Churchill*

Acknowledgement

I would like to thank my advisors Jean-Herve Tortai and Patrick Schiavone who took me on this journey and made their best to help me achieve the knowledge required to accomplish this task. I also would like to acknowledge their friendship that was important for coping with all changes: returning to the academia after five years working on the industry, moving to a different country with all the cultural and language differences, etc.

This work was developed as a CIFRE thesis inside Aselta Nanographics. I'm very thankful for this great team that supported me in several occasions and helped me develop this thesis work both with valuable ideas and their implementations. I would like to specially thank Clyde Browning who, from day one, shared his technical experience and his expat knowledge, both extremely valuable for me during this period. Moreover, I would like to thank Mohamed Saib for the all the research that we did together which is reflected in several different results presented on this thesis and the patents originated from it. Sébastien Bayle and Luc Martin were also really important during this period helping me understanding Inscale and Aselta all at once. The friendship and support from the entire Aselta team was remarkable. I would like to express my gratefulness for the discussions, the moments shared and encouragement they all gave me: Alexandre Chagoya, Alexis Blanchemain, Anne Adolphe, Bruno Gilles, Charles Tiphine, Christophe Guittard, Céline Tranquillin, Guillaume Renault, Hervé Guegan, Jean Mirallie, Julien Nicoulaud, Mathieu Dirrenberger, Matthieu Millequant, Nassim Halli, Paolo Petroni, Pascal Borot, Serdar Manakli, Sergei Postnikov, Stéphane Buisson, Thomas Quaglio, Vincent Annezo and Virginie Baticle. Thank you for all the cultural advices, French lessons, jokes, wine and cheese knowledge sharing, and basic information that made this period more meaningful and pleasant.

I would like to thank the Laboratoire des Technologies de la Microélectronique (LTM) for the opportunity of developing this thesis and for the collaboration with Sebastien Labau and Nader Jedidi. I would also like to thank the support from Jumana Boussey during this entire period.

During this period I have also cooperated with people from different Aselta partnerships. I've learned a lot from these interactions with Fraunhofer CNT and CEA LETI. I would like to specially acknowledge Kahn-Hoon Choi, Christophe Hohle, Michael Kaiser, Xaver Thurn, Manuela Gutsch and Katja Steidel, from CNT and Aurélien Fay, Jonathan Pradelles, Cyril Vannuffel, Martin Thornton, Stephanie Meunier, Jessy Bustos and Jerome Belledent from LETI. These collaborations enriched my thesis work and contributed significantly for improving my technical knowledge on this field which was not my own.

I also would like to thank the friends I made in France since our arrival. Especially I would like to thank Marie Browning and Justin Rouxel for their friendship and the nice conversations we had over these almost four years.

Moreover, I want to thank my mother Cristina, my brothers Rodrigo and Otávio, and my friend Regis for the support and understanding. They understood my absence in several important family moments. I also would like to thank how they always made me feel at home. I thank them for the visits both in Brazil and in France, for the trips we did together as well as the Skype calls that kept us close

together at any time. I would also like to thank my godsons Vitor and João Pedro. I hope I will be able to compensate the absence during this important part of their lives and Fernando, who I only could be with for two short months and knows me more as the “guy on the computer screen”.

Finally, I would like to thank my wife that took this adventure with me for sharing the burden and multiplying the joys. I thank her for moving to a distant country, learning its language and culture and facing the distance from friends and family with me. Without her love and support, I doubt that this mission could have being accomplished at all.

Abstract

Since the development of the first integrated circuit, the number of components fabricated in a chip continued to grow while the dimensions of each component continued to be reduced. For each new technology node proposed, the fabrication process had to cope with the increasing complexity of its scaling down. The lithography step is one of the most critical for miniaturization due to the tightened requirements in both precision and accuracy of the pattern dimension printed into the wafer. Current mass production lithography technique is optical lithography. This technology is facing its resolution limits and the industry is looking for new approaches, such as Multi-patterning (MP), EUV lithography, Direct Write (DW), Nanoimprint or Direct Self-Assembly (DSA). Although these alternatives present significant differences among each other, they all present something in common: they rely on e-beam writers at some point of their flow.

E-beam based lithography is subject to phenomena that impact resolution such as are electron scattering, fogging, acid diffusion, CMP loading, etc. The solution the industry adopted to address these effects is to predict and compensate for them. This correction requires predicting the effects, which is achieved through modeling. Hence the importance of developing accurate models for e-beam process.

In this thesis, the basic concepts involving modeling are presented. Topics such as data quality, model selection and model validation are introduced as tools for modeling of e-beam lithography. Moreover, the concepts of local and global sensitivity analysis were also presented. Different strategies of global sensitivity analysis were presented and discussed as well as one of the main aspects in its evaluation, which is the space sampling approach.

State-of-the-art strategies for todays and future lithography processes were presented and each of their main steps were described. First Principle models that explain the physics and chemistry of the most influential steps in the process resolution were also discussed. Moreover, general Compact models for predicting the results from e-beam lithography were also presented. Finally, some of the limitations of the current approach were described.

New compact models described as Point-Spread-Function (PSF) are proposed based on new distributions, such as Gamma and Voigt. Besides, a technique using Splines for describing a PSF is also proposed. Moreover, a flexible resist model able to integrate most of the observed behavior was also proposed, based on evaluating any pattern on the layout using metrics. Results using such method further improved the any of the PSF distribution approach on the critical features that were limiting the future technology nodes. Other specific models and strategies for describing and compensating for extreme-long-range effects is also proposed and described in this work.

The calibration layout is a key factor for providing the calibration algorithm with the experimental data necessary to determine the values of each of the parameters of the model. Several strategies from the literature were briefly described before introducing one of the main propositions of this thesis, which is employing variance-based global sensitivity analysis to

determine which patterns are more suitable to be used for calibration. A complete flow for selecting patterns for a calibration layout was presented. A study regarding the impact of process and metrology variability over the calibration result was presented, indicating the limits one may expect from the generated model according to the quality of the data used. Finally, techniques for assuring the quality of a model such as cross-validation were also presented and demonstrated in some real-life situations.

Résumé

Depuis l'apparition du premier circuit intégré, le nombre de composants constituant une puce électronique n'a cessé d'augmenter tandis que les dimensions des composants ont continuellement diminué. Pour chaque nouveau nœud technologique, les procédés de fabrication se sont complexifiés pour permettre cette réduction de taille. L'étape de lithographie est une des étapes la plus critique pour permettre la miniaturisation. La technique de lithographie qui permet la production en masse est la lithographie optique par projection. Néanmoins cette technologie approche de ses limites en résolution et l'industrie cherche de nouvelles techniques pour continuer à réduire la taille des composants. Les candidats sont l'écriture en plusieurs passes, la lithographie EUV, l'écriture directe, la nano-impression ou l'auto-organisation dirigée. Même si ces alternatives reposent sur des principes très différents, chacune a en commun l'utilisation de la lithographie électronique à un moment ou à un autre de leur réalisation.

La lithographie électronique est sujette à des phénomènes spécifiques qui impactent la résolution finale, tels la diffusion des électrons, le « fogging », la diffusion d'acide, la CMP etc… La solution choisie par l'industrie pour tenir compte de tous ces phénomènes est de les prévoir puis de les compenser. Cette correction nécessite de les prédire à l'aide de modélisation, la précision de ces modèles décrivant les procédés étant primordiale.

Dans cette thèse, les concepts de base permettant de développer un modèle sont présentés. L'évaluation de la qualité des données, la méthodologie de choix d'un modèle ainsi que la validation de ce model sont introduites. De plus, les concepts d'analyse de sensibilité locale et globale seront définis.

L'état de l'art des stratégies utilisées ou envisagées pour les procédés lithographiques actuels ou futurs sont énoncés, chacune des principales étapes lithographiques étant détaillée. Les modèles tenant compte de la physique et de la chimie impactant sur la résolution après écriture par e-beam sont étudiés. De plus, les modèles compacts permettant de prédire les résultats obtenus par e-beam seront détaillés, pour finalement décrire les limitations des stratégies actuelles.

De nouveaux modèles compactes sont proposés en introduisant de nouvelles familles de fonctions telles que les fonctions Gamma ou les fonctions de Voigt. De plus, l'utilisation des fonctions d'interpolations de type Spline sont également proposés. Un modèle résine d'utilisation souple a également été développé pour tenir compte de la plupart des comportements expérimentaux observés en évaluant les dimensions de motifs d'un dessin en utilisant des métriques appropriés. Les résultats obtenus en utilisant de telles méthodes montrent une amélioration de la précision de la modélisation, notamment en ce qui concerne les motifs critiques. D'autres modèles spécifiques permettant de décrire les effets d'extrême longue portée sont également décrits dans ce travail.

Le choix du jeu de motifs de calibration est critique pour permettre à l'algorithme de calibration d'obtenir des valeurs robustes des paramètres du modèle. Plusieurs stratégies utilisées dans la littérature sont brièvement décrites avant l'introduction d'une technique qui utilise l'analyse de sensibilité globale basée sur la variance afin de sélectionner les types de géométries optimales pour la calibration. Une stratégie permettant la sélection de ces motifs de calibration est détaillée. L'étude de l'impact du procédé et des incertitudes de mesures issue de la métrologie est également abordée, ce qui permet d'énoncer les limites à attendre du modèle sachant que les mesures peuvent être imprécises. Finalement, des techniques permettant de s'assurer de la qualité d'un modèle sont détaillées, telle l'utilisation de la validation croisée. La pertinence de ces techniques est démontrée pour quelques cas réel.

Summary

Introduction

"In the strict formulation of the law of causality—if we know the present, we can calculate the future—it is not the conclusion that is wrong but the premise."

- *Werner Heisenberg*

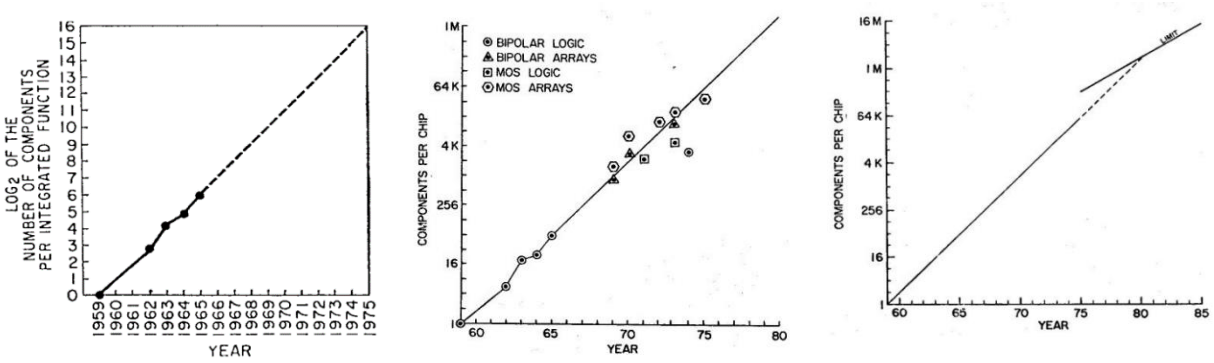
Introduction

Since the development of the first integrated circuit, the number of components in a chip continued to grow and the dimensions of each component continued to be reduced. These two effects combined give rise to impressive improvements in system performance. Reducing the size of transistors improves their performance [DENNARD 1974]. Moreover, adding more transistors to the same circuit increases its capacity for arithmetic and logic operations as well as its memory. As the integrated circuits became more powerful and versatile, more sectors of the industry started employing them in order to improve their own products. Nowadays, even equipment as simple as coffee machines and dishwashers may present a few integrated circuits, while complex ones, such as cars and airplanes, may present from hundreds to thousands of chips.

The economic model of this apparently ever expanding market relies on achieving and maintaining a given pace when it comes to higher circuit integration. This pace is well known in the industry as the Moore's law. In its most common definition, Moore's law predicts that the number of transistors in a chip would double every two years [MOORE 1975]. In another formulation it says that computer performance would double every eighteen months. This last formulation, although very common, was never stated by Moore [MOORE 1995]. Eventually, many other predictions stating that something would double every eighteen or twenty four months have been called Moore's law as well.

In his original formulation [MOORE 1965], the Moore's law predicted that the number of components in an integrated circuit would double every year. Figure 1 (a) shows the graph used originally by Moore indicating the number of components in a chip from 1959 to 1965 and the estimation for the next ten years (according to the Moore's law). In 1975, Moore changed his law [MOORE 1975] for expecting the doubling every two years, instead of every year, as he expected that most of the innovating possibilities on how the relative component distribution could be improved were already made, as shown in Figure 1 (c). It is interesting to remark that in the ten years that separate both publications, the number of components roughly double every year, as presented in Figure 1 (b). Moreover, right after the second publication, the number of components doubled every two years, exactly as predicted again by the 1975 Moore's law.

Therefore, this behavior allowed the industry to sell more powerful chips for the same price or to reduce the price for the same chip. This business model has being working very well while the fabrication technology allowed to continually scale down the circuit components. Since the entire economic model of this industry relies in Moore's law, it may even be seen as a self-fulfilling prophecy.



(a) (b) (c)

Figure 1 (a) presents the original prediction from Moore's work in 1965 [Moore 1965]. (b) shows the data until 1975, indicating that the first prediction was right [Moore 1975] and (c) shows the inflection point where Moore predicted the law change from doubling every year to doubling every two years [Moore 1975].

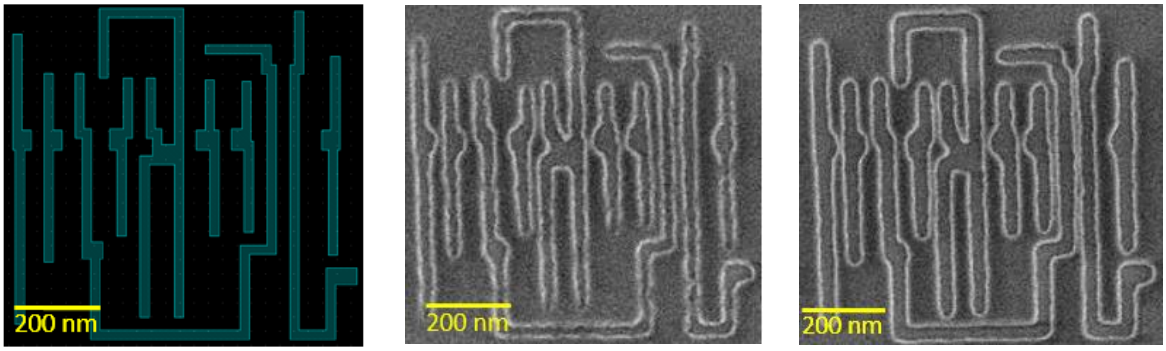
For each new technology node proposed by the industry, the complexity of the fabrication process increases to deal with the increasing constraints due to its scaling down. Hence, each new process of fabrication of integrated circuits requires a larger number of steps, which are all being constantly improved in order to achieve the requirements for component scaling down. Among these steps, lithography is one of the most critical for miniaturization due to the tightened requirements in both precision and accuracy of the pattern dimension printed into the wafer. Therefore, the lithography process is one of the pillars of the entire business model of the semiconductor industry due to its direct relation with achieving the scaling down required by the Moore's law.

There are several different lithography techniques. The technique used in mass production is optical lithography or photolithography. It is favored by the industry because it presents the throughput and resolution required by today's technology at the best cost. This technique is based on using a photomask, which is a transparent plate with opacities that allows writing the circuit by either letting the light pass or blocking it, patterning the design in a photoresist. The quality of the patterns present in the photomask is directly proportional to the quality of the photolithography resolution. In integrated circuit fabrication processes several masks are used, at least one for each layer of the circuit. For the most critical layers (i.e. the ones presenting very small features, high density or requiring highly controlled dimensions), the masks are written using an electron beam (e-beam) machine, which presents high accuracy and precision and the resolution required for writing such critical patterns.

Nevertheless, nowadays optical lithography is facing resolution limitations and, in order to continue to shrink components dimensions, the industry is looking for a better solution. Some of the candidate solutions are Multi-patterning (MP), Extreme-Ultra-Violet (EUV) lithography, Direct Write (DW), Nano-imprint and Direct Self-Assembly (DSA). Although these alternatives present significant differences among each other, they all rely on e-beam writers at some point of their flow. This makes of e-beam machines a meeting point for the future of lithography.

E-beam writers present high accuracy, but they are subject to phenomena that limit their resolution as well. The beam presents a diameter which is not negligible; some of the electrons

scatter while interacting with the material and may expose regions that were supposed to be left unexposed. Electrons leaving the surface of the material may be reflected back to any other spot and so on. These effects, combined with other factors such as resist development and plasma etching characteristics directly impacts the resolution of the patterns written by e-beam machines. Besides the natural approach of working in improving these machines, the solution the industry adopted to address these effects is to predict and compensate for them. Traditionally, this compensation is called Proximity Effect Correction (PEC), although it may account for several different factors. Nonetheless, this correction requires predicting the effects, which is achieved through modeling. Figure 2 illustrates the importance of developing accurate models for e-beam processes. Figure 2 (a) shows a pattern to be printed, while in Figure 2 (b) one may see a low quality result due to poor compensation and Figure 2 (c) shows a much better result, after proper compensation.



(a) (b) (c)

Figure 2 (a) presents a pattern to be printed, (b) a SEM picture of the same layout after exposure based on compensation of a poor model and (c) a SEM picture of the same layout after compensation using a better model.

E-beam modeling has being developed together with the first e-beam machines. However, as technology advances, the requirements of modeling accuracy are getting tighter. To respond to it, the models are also becoming more complex. The model calibration effort is therefore increasing and the impact of every small change in the writing process has to be directly reflected in the model being used. In this new scenario, new models and strategies for calibrating them are continually being proposed.

The main goal of this thesis is to provide a complete flow on how to create an accurate and useful model for e-beam lithography both for simulation and for correction. In the process of describing such flow, most of the key aspects that contribute to the e-beam processes are described. Moreover, some of the strategies for determining the best model for a given process and for selecting the best test patterns to be used in an experimental calibration are also described. State-of-the-art e-beam models are discussed and their weaknesses exposed to finally present new model structures that enables continue to advance to future technology nodes using current process flows.

Chapter I presents the basic concepts of the semiconductor industry, mainly focusing on integrated circuit fabrication and the aspects of the lithography. The most prominent candidate

lithography techniques to be used in future technology nodes are also discussed and their weakness and strengths are pointed out.

In chapter II the aspects of producing, validating and comparing models are presented. Basics on data analysis are also discussed to be employed in the evaluation of the model calibration process. Special attention is given to the criteria to evaluate the quality of a model, presenting different statistical algorithms for performing biased, unbiased and robust evaluations.

The chapter III presents the concepts of Sensitivity Analysis in both their local and global forms. More emphasis is given to Variance-based Sensitivity Analysis since it is the more suitable approach for the applications required in this thesis. Sampling techniques and evaluation criteria are also presented and discussed. Although general information are presented, the focus is on the aspects necessary for the further development of the work of this thesis.

Chapter IV describes the characteristics and issues of e-beam lithography and stresses the necessity of modeling them. Moreover, strategies for modeling e-beam lithography both using first principle models and compact ones are discussed and new compact models are proposed in order to obtain higher precision than the state-of-the-art of the industry.

The chapter V focus on the model calibration, either relying in first principle model results or, especially, relying in experimental data, which is becoming the industry standard. Moreover, the aspects related to generating and verifying a design to be used in the experimental calibration procedure are discussed. The strategy to ensure a calibration design that is suitable to the aimed process models and is robust to the variability of the process is detailed. Finally, the optimum strategies to validate the generated lithography model are presented.

In chapter VI another application developed in the period of this thesis is presented. It consist of an strategy to compensate for extreme-long-range effects on large data, such as processing the entire reticle (for mask writing) or the entire wafer (for direct write). This is performed based on a density evaluation and performing the convolution of the generated model over the density map instead of over the original data.

Chapter VII presents the results of employing the techniques discussed in this thesis for e-beam modeling. The results of model calibration exercises over the results of first principle models and experimental data is presented. Moreover, some specific topics regarding sensitivity analysis were evaluated and are discussed as well. The evaluation of a model using cross-validation is also presented, as well as a study showing the impact of process variability over the calibration result. Finally, a result of the evaluation of the extreme-long-range impact over the direct-write exposure.

Finally, the last chapter of this work presents the conclusions and perspectives for continue to improve the performance of e-beam exposures using prediction and compensation in order to address future technology nodes.

Chapter I. Basic Concepts

"The future influence the present just as much as the past."

- *Friedrich Nietzsche*

1. Integrated Circuit Fabrication

The fabrication of integrated circuits (ICs) is a multiple-step sequence during which circuits are created on a wafer made of semiconducting material. A typical state-of-the-art wafer is a circular slice about 0.75mm thick and 300mm in diameter. The most frequently used wafer is silicon, but some other approaches exist, notably the Silicon on insulator (SOI). The steps for producing the circuit on a wafer are divided into Front-End-of-Line (FEOL) and Back-End-of-Line (BEOL). FEOL is where the individual devices – such as the transistors – are patterned in the semiconductor. The BEOL is responsible for interconnecting the devices created during the FEOL. Generally it begins when the first layer of metal is deposited on the wafer. BEOL includes contacts, insulating layers, metal layers and goes up to the bonding sites for packaging.

During both FEOL and BEOL, each step of adding a new material layer requires at least three steps, which are the deposition, the patterning and the removal. Deposition consists in growing, coating or transferring a material onto the wafer by technologies such as physical vapor deposition (PVD), chemical vapor deposition (CVD), among others. Patterning consists in shaping or altering deposited materials. The removal step consist in removing material from the wafer, by processes such as etching or chemical-mechanical planarization (CMP). After these three steps, a resist strip is performed to prepare the material for the next process. Figure 3 illustrates the steps of this flow.

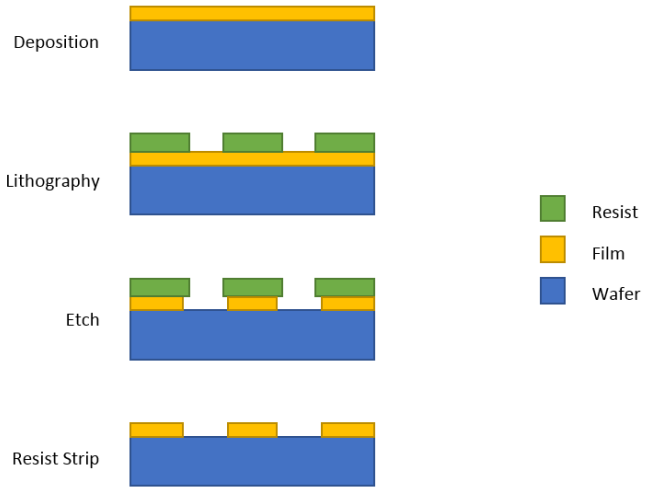


Figure 3. Microelectronic process for one layer patterning.

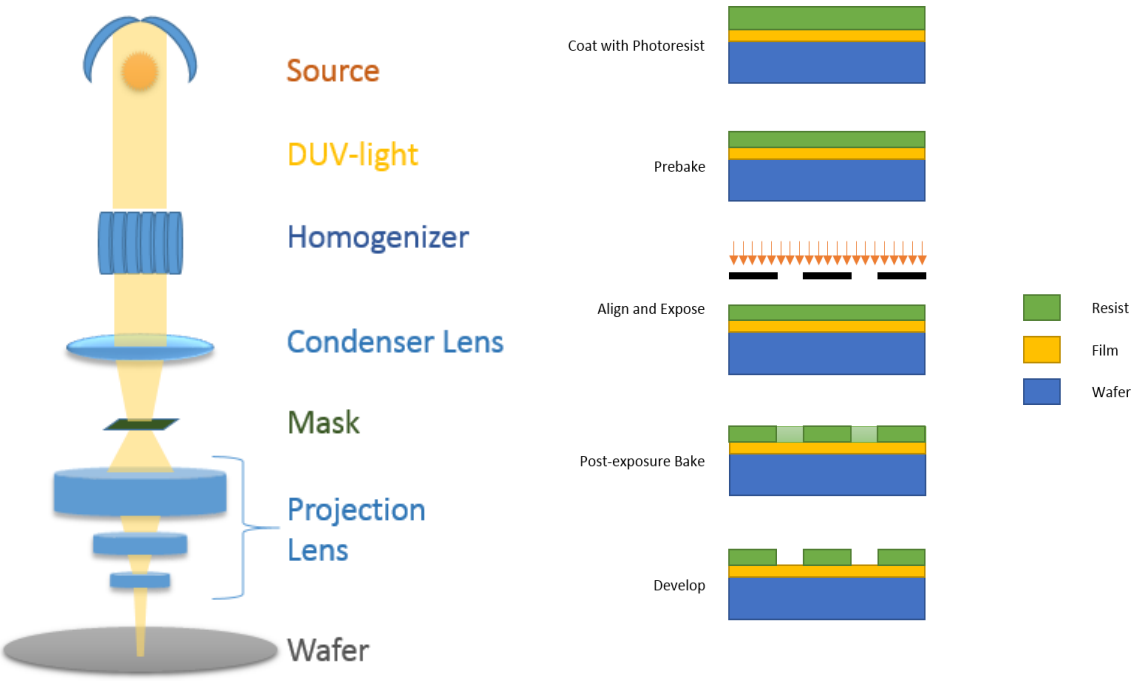
1.1. Lithography Process

As mentioned in the previous section, a chip is composed of several layers and, for patterning each layer, it is necessary to perform one or more lithography steps. There are several different strategies to transfer the patterns of a circuit into the silicon wafer. Nonetheless, the industry current standard is photolithography.

1.1.1. Photolithography

Photolithography is the process where a light-sensitive polymer, called photoresist, is exposed and developed to form 3D patterns on the substrate. Ideally, the final image found at the photoresist is identical to the intended pattern. The local exposure of the photosensitive polymer is achieved through a photomask. This mask is a quartz plate presenting opaque and transparent features either blocking the light or allowing it to shine through. The intended patterns to be exposed in the substrate are, therefore, first drawn on the mask and, then developed on the substrate. The process of writing the intended patterns on the mask is very similar to a lithography process itself for a single layer and usually relies on laser writers or electron-beam writers (details on electron-beam lithography are presented further on this chapter). Usually patterns present in a mask present a scale factor in comparison to the final dimensions expected on the wafer. In state-of-the-art industrial production, patterns on a mask are four times bigger than the desired dimensions on wafer.

Figure 4 (a) represents a simplified scheme of a photolithography exposure tool while (b) shows the five main steps of the lithography process, from coating the resist to the development of the intended pattern on the wafer.



(a) (b)

Figure 4. (a) Simplified scheme of a photolithography exposure tool while (b) shows the five main steps of the lithography process.

The resolution of this system is limited by the wavelength of the light used and the ability of the projection lens system to capture enough diffraction orders after passing through the mask. Lithography systems have progressed from blue wavelengths, 436nm, to ultraviolet (UV),

365nm, and to deep ultraviolet (DUV), 248nm, and finally state-of-the-art photolithography tools employ (DUV) light with a wavelength of 193nm combined with immersion fluid.

The minimum feature size that an optical projection lithography system can achieve may be approximated by the so called Rayleigh formula (1).

$$CD = \frac{k_1 \lambda}{NA} \quad (1)$$

Where CD is the critical dimension, or the minimum size of feature that may be exposed by the tool, λ is the wavelength of the light employed by the tool, NA is the numerical aperture of the lens from the wafer perspective and k_1 is a coefficient that lumps several process-related factors and is larger or equal to 0.25.

The numerical aperture is computed as shown in (2) and it is a convenient way to describe the size of the lens aperture in the lithography system. NA value for optical lithography DUV 193i is 1.35 while for extreme ultraviolet (EUV) it is 0.25 or 0.33 [ITRS, 2013].

$$NA = n \sin \theta_{max} \quad (2)$$

Where θ_{max} is the maximum half-angle of the diffracted light that is able to exit the objective lens while n is the refractive index of the medium between the wafer and the lens.

According to this equation, the CD may be decreased by increasing the numerical aperture and/or by reducing the wavelength. However, another major concern of this system is the depth of focus (DoF), which is the tolerance of height placement in the image plane in relation to the lens. The depth of focus can be approximated by (3).

$$DoF = \frac{k_2 \lambda}{NA^2} \quad (3)$$

Where k_2 is another coefficient that lumps several process-related factors.

Photolithography end has been predicted in many occasions but has continually advanced in order to address pattern sizes that fulfilled ITRS targets to meet Moore's law. For instance, years ago it was supposed to be impossible to write features smaller than 1µm by an optical strategy. More recently, another barrier was broken when features smaller than the exposure wavelength were also printed successfully using photolithography. Nowadays, once again, photolithography seems to have reach a hard point. The project to increase the resolution limits by using 157nm wavelength failed due to technical and economic problems. Last significant improvement in this system was the adoption of the immersion technique for 193nm steppers.

Immersion lithography is a technique that replaces the air in the gap between the final lens and the wafer surface by a liquid with a refractive index greater than one. Increasing n increases the numerical aperture NA, as seen in (2).The improvement in resolution is by a factor of the refractive index and, today, the industry employs highly purified water, whose index is 1.44 at a wavelength of 193nm. Moreover, this approach may be used to improve the DoF by at least

the refractive index of the fluid and by larger factors as the pitches become smaller [MACK, 2007].

Therefore, the industry continues to look for solutions to achieve smaller technology nodes. Currently solutions are benchmarked to manufacture the 14nm node, which is likely to require a new breakthrough on the lithography process. On the next section some candidate solutions for becoming the industry main stream lithography technique are presented.

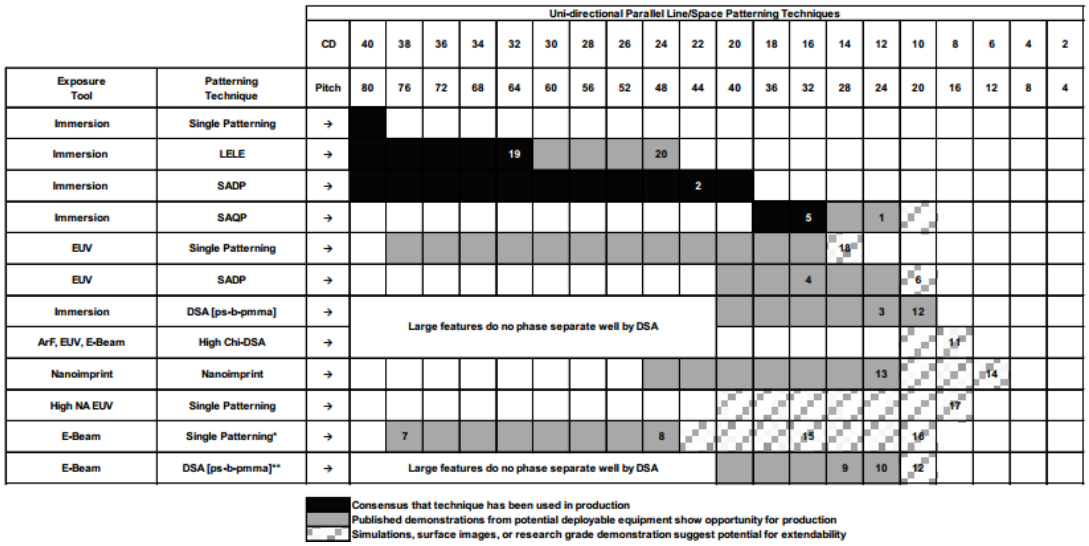
1.2. Candidate Lithography Processes

The industry have been improving the lithography processes by reducing the wavelength. This trend stopped at 193nm. The strategy of choice to continue evolving was the 157nm wavelength, which appeared in both ITRS and Intel [SILVERMAN, 2002] roadmaps as the replacer of the 193nm. However, in spite of the large investment and the advances achieved, it was eventually abandoned for good in favor of the immersion lithography approach (193i optical lithography).

When the 193i reached its limits, the industry continued looking for a long term solution. The approach of choice for a long time has been the Extreme-ultraviolet (EUV) lithography at 13.5nm wavelength. This approach has being in roadmaps as a future technology even before the use of immersion [SILVERMAN, 2002] but is still lacking some advances to overcome its current limitations. Meanwhile, other approaches were proposed, such as the Direct Write (DW) using electron-beam (e-beam) machines, nanoimprint and, more recently, multiple-patterning and direct-self-assembly (DSA). Yet the industry seems to be already employing multiple-patterning to achieve the 28nm and 22nm technology nodes [ITRS 2013].

A summary of each lithography strategy and its capabilities for reaching the CD and pitch values are given in Table 1 [ITRS 2013]. The black cells indicate where there are published papers indicating the use of the technique in production, grey cells and cross hatched ones indicate where techniques have been demonstrated to be capable of smaller resolutions.

Table 1. Demonstrated Line and Space Resolution of Potential New Patterning Techniques (from Figure LITH1A of ITRS 2013)



Observe that Double Patterning (Immersion SADP) and Quadruple Patterning (Immersion SAQP) are in production for achieving CDs of 18nm and 16nm. While the 14nm node is still only demonstrated as possible but there are several approaches that may reach such resolution (Quadruple Patterning, EUV with Double Patterning, Nanoimprint or approaches based on DSA).

1.2.1. Extreme-ultra-violet (EUV) Lithography

Photolithography has been advancing in the past decades by reducing its wavelength in order to improve its resolution. As the wavelength becomes smaller, it is difficult to obtain non-absorbing materials to be employed as optics of the lithography systems. Most of the light gets absorbed by the lenses. Nonetheless, a strategy was proposed in order to drastically reduce the wavelength, known as Extreme-ultra-violet (EUV) lithography [BAKSHI, 2009].

EUV wavelength is around 13nm, for which no set of transparent masks can be obtained. Therefore, EUV lithography relies in reflection through mirrors in order to draw the images on the wafer. Reflection occurs at interfaces between materials of different indices of refraction. The larger the difference of the indices of these two materials, the greater the reflectivity of their interface. However, at wavelengths such as the ones of EUV, all materials present indices of refraction around 1. Thus, it is difficult to create a highly reflective interface. Nonetheless, by using multilayers alternating high-Z (chemical elements with a high atomic number) and low-Z (chemical elements with a low atomic number) materials, it is possible to obtain mirrors with reflectivity in the range of 60-70%. Basically, it is the accumulation of interfaces of small reflectivity that allows reaching reasonable total reflectivity, if a sufficient number of layers is employed.

Besides the criteria of being high-Z and low-Z materials, the chosen materials must also be weak absorbers of EUV light. This is because the light must be able to penetrate to the lower layers of the film stack. Some stacks have been identified and the industry tends to use pairs of Molibdenum (Mo, Z = 42) and Silicon (Si, Z = 14) for the multilayer reflector at 13.5nm. This stack is usually called the MoSi stack.

Another significant change from 193i lithography is the change in the absorber material used in the mask. For EUV, Cr is no longer sufficient and heavier materials, such as the Ta, are used instead. An example of the difference in the materials stack for a photolithography mask and a EUV mask is presented in Figure 5. Observe that EUV requires more layers and heavier materials, which have impact on the mask writing procedure, topic that will be further explored later in this thesis.

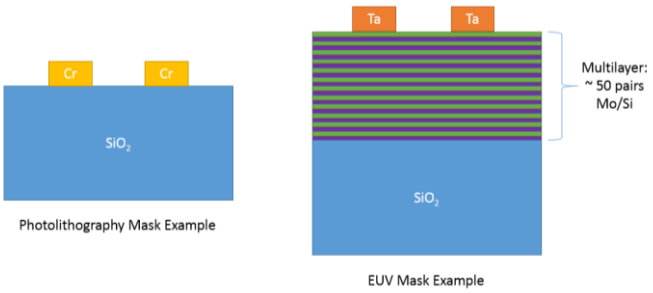


Figure 5. An example of the stack of materials for a photolithography mask and a EUV mask.

Using mirrors instead of lenses requires changes in the exposure tool as well. Figure 6 shows a schematic of a EUV tool with its sets of mirrors and the reflecting mask.

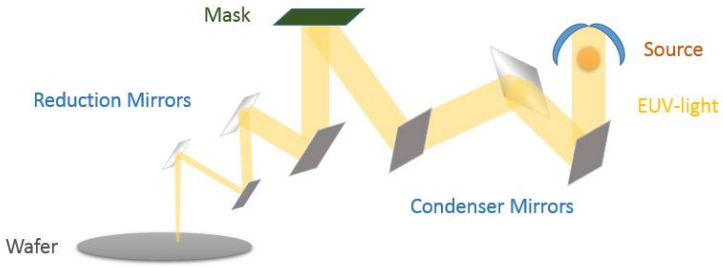


Figure 6. A simplified representation of the EUV exposure system.

However, since at its best each mirror (and the mask itself) absorbs 30% of the energy of the EUV light, the original energy at the source must be extremely high in order to compensate those losses. If the amount of energy on the wafer is low, the exposure must continue for a longer period, increasing significantly the writing time of the wafer. Current requirement is 150W of EUV light at intermediate focus [BAKSHI, 2009]. This is being one of the most significant limitations to achieve production level in the industry using EUV.

Finally, EUV lithography tools must be under vacuum in order to avoid photons absorption. Hence a new technical difficulty appears during exposure of photoresist due to resist outgassing and molecule cracking under EUV. This cracking induces amorphous deposition onto mirrors, reducing significantly the reflectivity continuously.

Since EUV failed to be ready to replace 193i in time for the technology nodes using pitch below 80nm, the industry found a fallback plan in using multiple-patterning as a strategy to relax some of the constraints of 193i optical lithography. This is the topic of the next section.

1.2.2. Multiple-patterning

The physical limitation of the optical lithography is not the size of the features alone but the association of this dimension and the space close to it. This constraint is usually represented by the notion of pitch. Since, after the Rayleigh criterion, the resolution limit of standard 193i is close to be the 90nm pitch, the industry found in multiple-pattering the strategy to continue to

operate with most of the same process flow of the 193i and achieve technology nodes of 28nm and below [FONSECA, 2009][OWA, 2014] [MAILFERT, 2014].

There are several variations of this approach, either relying in a single exposure (and, therefore, in a dual-tone photoresist or a dual-tone development, for instance) or in multiple exposures (using a single or multiple resists). Although using a single exposure has the advantage of not increasing the number of masks required for drawing the intended patterns, the complexity added to the process by modifying the resists or developments makes them hardly viable for industrial applications.

Multiple exposure is based on the strategy of separating a layout into different masks in order to reduce the feature density on each exposure. The decision of which patterns should be drawn in which exposure, usually called split, is usually performed automatically by graph coloring algorithms [MAILFERT, 2014]. Relaxing the pitch allows to print correctly very small features. Figure 7 shows a representation of patterns to be exposed and the points where the pitch should be relaxed (distance violation). It also shows the coloring result and the two resulting masks (first mask with the features from coloring group A and second mask with the ones from coloring group B) for a double exposure procedure.

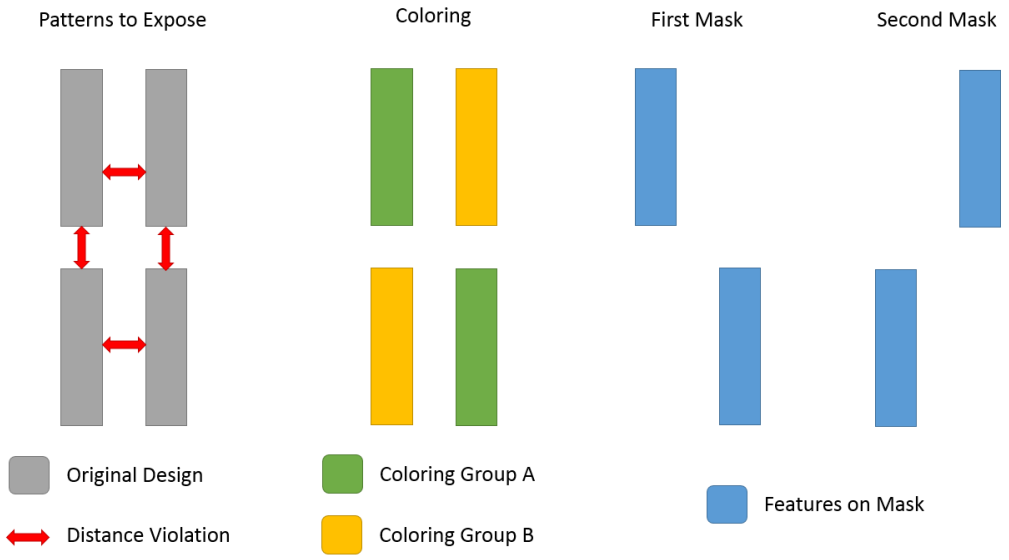


Figure 7. A representation of the partition of the patterns to expose in order to prevent the violation of the minimal distances. The coloring and the resulting two masks.

However, not all patterns are printable using a double exposure approach as presented in Figure 7. In some configurations more than two colors are necessary, as the example shown in Figure 8.

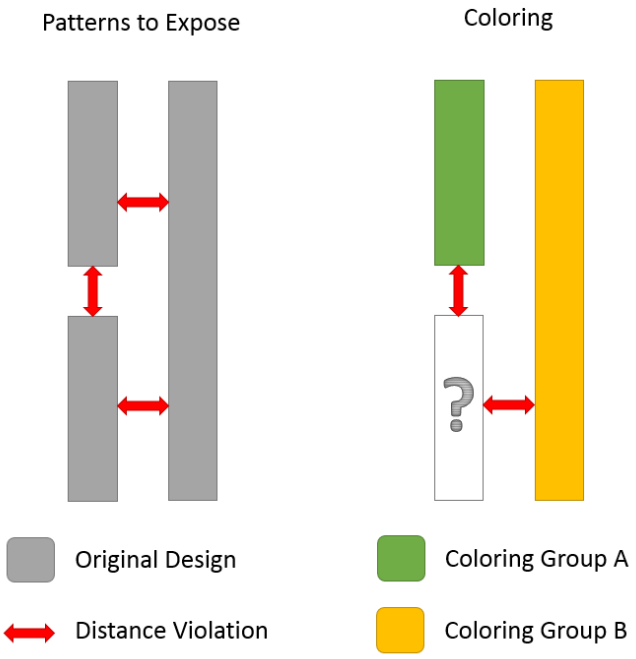


Figure 8. A representation of the partition of the patterns to expose in order to prevent the violation of the minimal distances and the impossibility to define the two colors and remove all violations.

The issues of coloring such as the one presented in Figure 8 may still be addressed by only two layers if one employs a pattern splitting technique, for instance. However, dividing lines increase the registration criticality and also works as a resolution limitation of this approach.

Although quadruple exposures [OWA, 2014] are a reality and even more exposures are feasible in order to further relax the half-pitch constraint, this approach is not viewed as an ideal candidate solution by the industry, (at least not for several layers). The reason for that is mostly related to the cost of producing so many masks and having so many process steps (after all, the Moore's law is about economics, not physics). Moreover, challenges of alignment as the number of exposures increases drastically. Due to these limitations, the industry is still looking for other solutions in order to continue evolving from one technology node to the next. Since EUV is still late, in the past few years another solution has increase attention due to its small feature sizes and the fast advances obtained by research groups. It is known as Directed self-assembly (DSA).

1.2.3. Directed self-assembly (DSA)

The Directed self-assembly (DSA) approach relies on inducing alignment in sets of block copolymers to organize themselves in a directed way in order to create regular structures [TSAI, 2013][JEONG, 2013]. The base of DSA are the block copolymers, which present two different polymers in blocks instead of randomly distributed in a chain, as represented in Figure 9(a). According to the relative distribution of each polymer, the way they interact may turn in different organized structures, as shown in Figure 9(b).

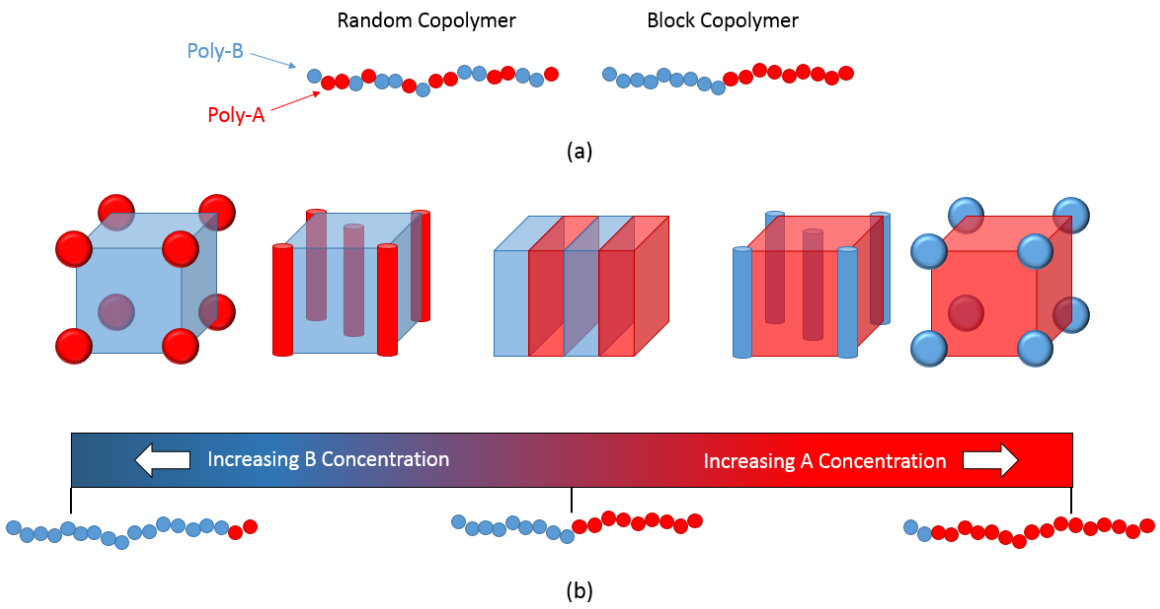
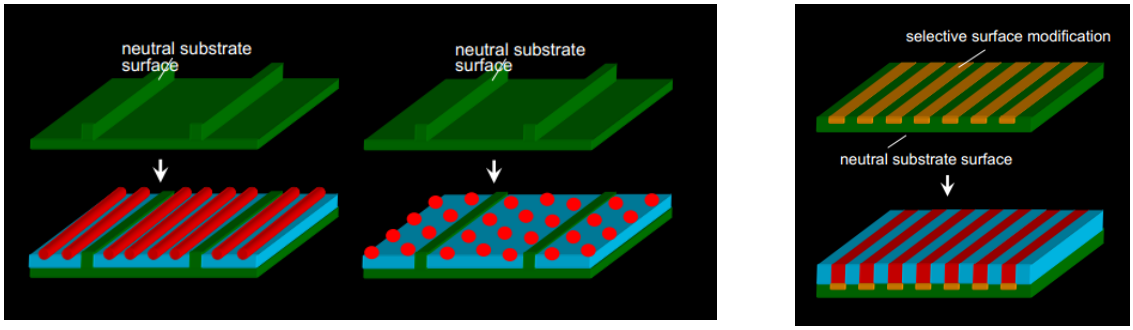


Figure 9. Basics on copolymers for DSA. (a) shows the difference of a random copolymer and a block one while (b) shows the difference in concentrations.

The dimensions in this approach are controlled by the molecular size and not by any diffraction parameter. Nonetheless, the ability to control the way how the molecules will place themselves is delicate. There are two main approaches to control the orientation of the molecules, namely graphoepitaxy and chemical epitaxy. The first relies on pre-existing topographic patterns generated by a standard lithography process with relaxed constraints while the second relies on surface patterns presenting a selective surface modification. Figure 10 shows an illustration of the difference between (a) the graphoepitaxy and (b) the chemical epitaxy approaches.



(a) (b)

Figure 10. Schemes for (a) graphoepitaxy and (b) chemical epitaxy.

Independently of the approach used, if the procedure is well succeeded the result is a very regular periodic structure. In order to convert the generated patterns into portions of real circuits it is necessary to remove some of the structures by using another lithography step. This step present more relaxed terms of density than if it was employed directly to generate the features itself, but still requires high accuracy and small dimension. This final step is usually performed using either photolithography employing a cut-mask or using electron-beam lithography. Figure 11 illustrates this process.

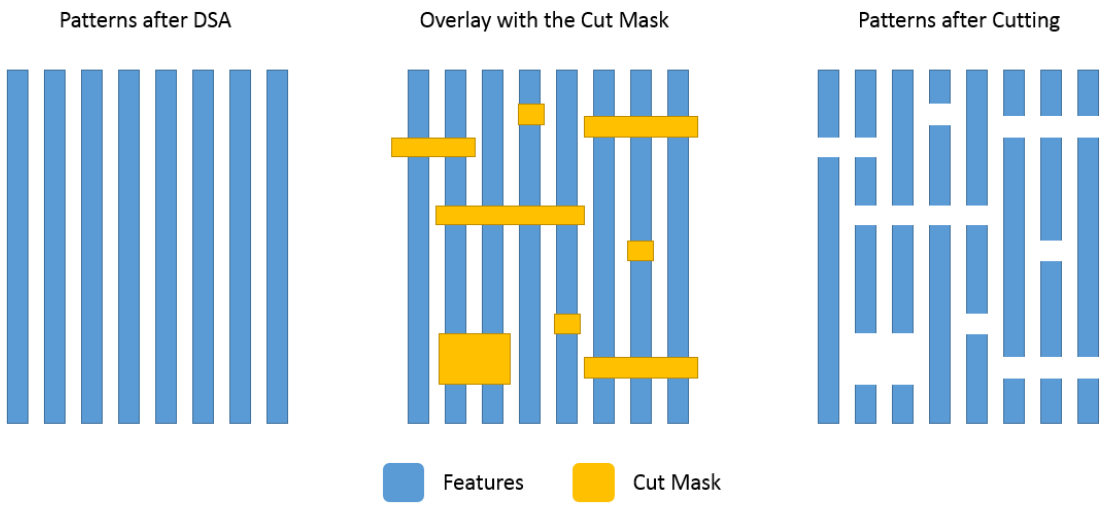


Figure 11. Example on how a procedure to remove some of the structures from the DSA in order to form the intended patterns.

Although DSA is a very promising approach and the industry is paying closer and closer attention to its development, it is still far from being a mature approach to be employed for future technology nodes. One of the main issues of this approach is the defectivity issue, which affects drastically the yield of this kind of lithography. In this sense the industry also looks for the direct-write approach using e-beam machines, which is the topic of the next section.

1.2.4. Electron-beam lithography

Electron-beam (e-beam) lithography has been known since the sixties. Its origins report to microscopy in which, by using electron sensitive resists, it was possible to employ the focused beam of electrons to draw custom shapes. In opposition to what is observed in the photolithography, most of the e-beam lithography techniques do not require the use of masks, being, on these cases, known either as maskless lithography or as direct write lithography [LEE, 1981][CRAIGHEAD, 1983].

The interest in employing electrons instead of photons to perform exposure comes from its high resolution. While photons wavelength for lithography are stalled at 193nm, the electron wavelength is of around 0.05 Å (or 0.005nm) at 50keV, disregarding the relativistic effects [MARTIN, 2011]. This very small wavelength makes the diffraction effects negligible.

1.2.4.1. Gaussian electron-beam lithography

The first and simplest strategy of e-beam lithography is based on Gaussian beam. A Gaussian e-beam tool usually employs a heated lanthanum hexaboride (LaB6) submitted to a very high voltage (several thousands of volts) as electron source. A column composed of several electromagnetic lenses condenses the electrons into a beam and allows it to focus into a given point of the wafer. Other lenses allow the beam to be deflected in order to draw the intended patterns. Figure 12 (a) shows a simplified example of such e-beam machine. This approach is called Gaussian beam due to the

scattering effect of the beam that distributes the incident electrons in a Gaussian probability around the focal point.

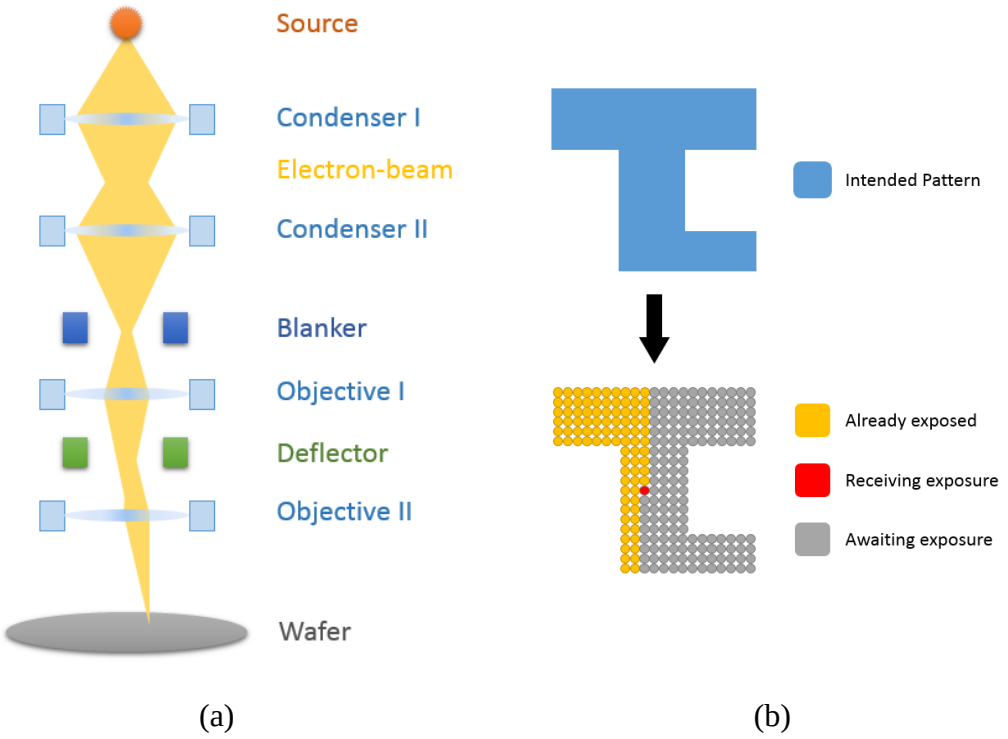


Figure 12. (a) A simplified representation of the Gaussian e-beam column and (b) The intended pattern being exposed by a Gaussian e-beam.

Using this strategy, the beam diameter may be smaller than 10nm [DIAL, 1998], [KURIHARA, 1995], [YAMAZAKI, 2000]. This dimension is inferior to the smaller features aimed for the industry in today's and near future technologies.

As the generated beam has a fixed shape and dimension, the intended patterns to be exposed are divided into very small features in a process called pixelization. In opposition to the photolithography, which expose an entire mask at once, the Gaussian beam lithography exposes one pixel at a time. This sequence of exposure is represented in Figure 12 (b). Even if it could be possible to increase substantially the time for exposure each pixel, the final exposure time using a Gaussian e-beam approach would still be way much slower compared to the photolithography. This is why this approach is not used in mass production of ICs and is mostly restricted to research.

Nonetheless, the electron lithography presented a series of advantages. One of them is the simple fact that a mask is not necessary to control the exposure so all the cost and challenges related to it are avoided. Instead, a software controls the exposure based on an input file. It also gives flexibility to the system as the files can be changed easily in contrast of a mask that, once printed cannot be modified.

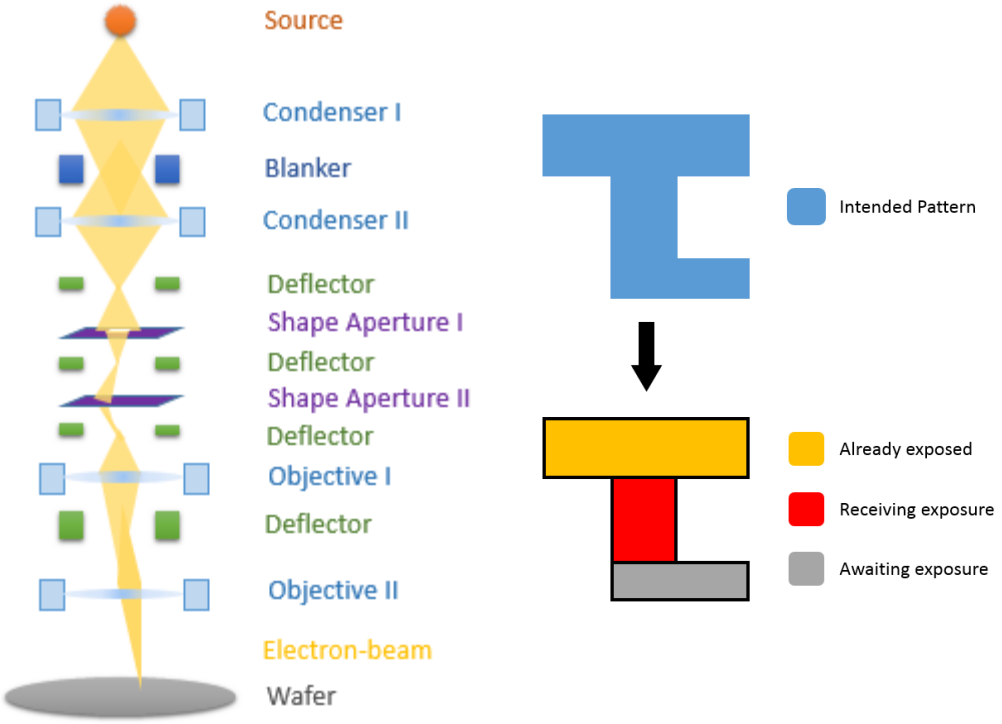
In order to improve the throughput of the Gaussian e-beam strategy, a new type of electron exposure was proposed, called shaped electron-beam lithography.

1.2.4.2. Shaped electron-beam lithography

One of the approaches proposed by the industry to improve the throughput of the electron lithography was the shaped electron-beam [RISHTON, 1999]. This approach relies on shaping the beam (hence its name) to cover a given geometric area on the target. In order to produce the shapes, the machine structure is very similar to the one used for the Gaussian beam but with the addition of shape aperture plates that control and give the desired geometry to the beam. A representation of the column is presented in Figure 13(a).

The geometry of the beam may vary in shape and size according to the ones available by the combination of the aperture plates present in the machine but, in general, may cover hundreds of points drawn by the Gaussian beam all at once. A scheme of a sequence of printing a pattern is represented in Figure 13(b). Observe that the intended pattern is exposed in three shots instead of the large number of dots presented in a Gaussian beam exposure (as shown in Figure 12 (b)).

Although considerably improving the exposure time of the Gaussian beam approach, this method is not fast enough to achieve the throughput required by IC mass production, which is usually stated as 100 wafers per hour (WPH). Current throughput of state-of-the-art shape beam is around 0.1 WPH [PFEIFFER, 2010]. Although its speed is a thousand times slower than the requirements for direct write, it is sufficient to be used as a method for mask writing for the photolithography. For this reason, the shaped e-beam lithography is the tool employed by the industry to write masks for 193i lithography, for EUV lithography and in all strategies of multiple patterning. However, as the technology advances, the structures become more complex due to more complex compensations for light diffraction and the mask writing time is becoming a concern for using shaped e-beam lithography.



(a) (b)

Figure 13. (a) A simplified representation of the shaped e-beam column and (b) The intended pattern being exposed by a Gaussian e-beam.

In order to further improve the throughput of electron-beam lithography the industry proposed to build arrays or matrices of beams (or shaped-beams) in order to expose a larger surface at once on the wafer. This approach is known as multi-beam and is discussed in the next section.

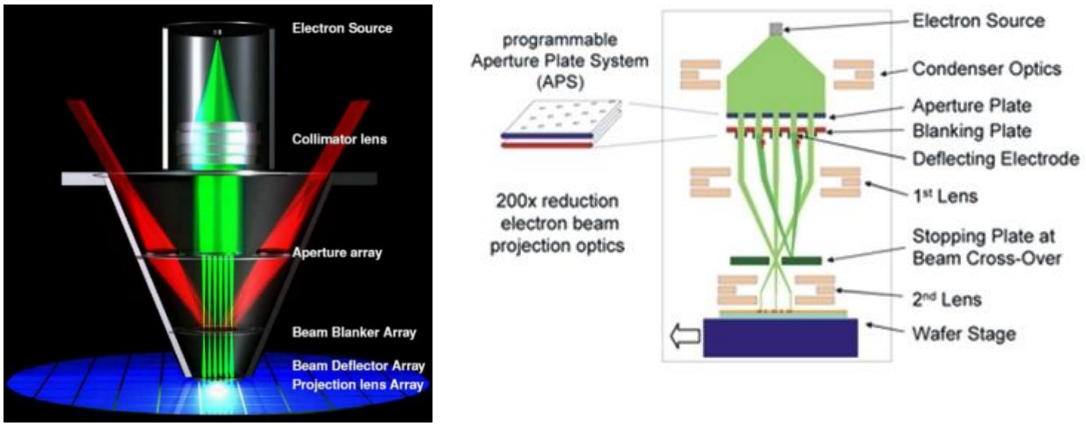
1.2.4.3. Multi-beam

To address the throughput limitations of e-beam writers the industry proposed recently the use of multiple beams in parallel, in a strategy known as multi-beam. There are two main variants of this approach, one presenting several beams in the same column and the other presenting several columns with one beam per column. The beams used can be either Gaussian (IMS Nanofabrication and MAPPER lithography) or Shaped (Vistec), the latter achieving little commercial success as of today. Another important difference among the strategies is the energy employed. While most of the approaches rely on high energy (around 50keV or 100keV), MAPPER adopted a "low" energy approach of 5keV. In any case, the throughput of these machines are increased by the parallel writing of all the beams, which may be used either for writing masks or even proposing to reach the performance required for direct write.

Generally speaking, the strategy splits a single beam in a large number (reaching thousands or even hundreds of thousands) of smaller beams. The number of beams vary from one implementation to another, as MAPPER FLX1200 announces 13,260 electron

beams in its production machine [BRANDT, 2014] while IMS Nanofabrication mention 262,144 programmable beams in their eMET [PLATZGUMMER, 2013].

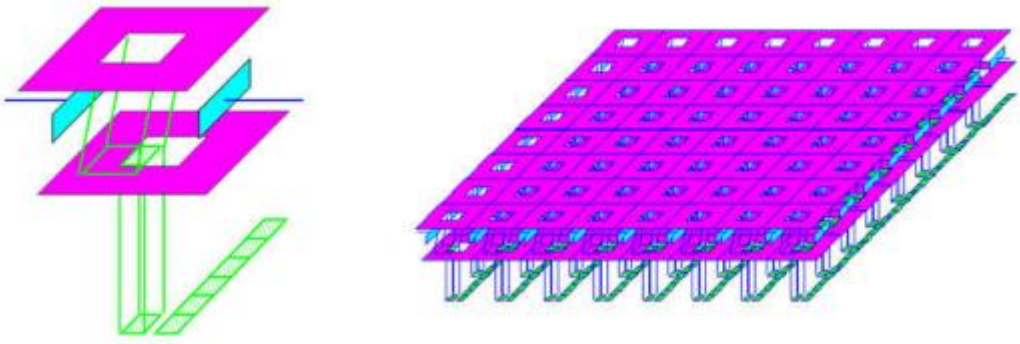
These beams scan the entire surface of the wafer or mask by the movements of the stage, while being turned on or off according to the requirements of the intended patterns. Since the entire surface is scanned independently if there are patterns to expose or not, the exposure time is no longer a function of the complexity of the patterns being exposed, qualifying this approach for complex mask writing for future nodes. Figure 14 shows the schematics of the tools proposed by (a) MAPPER and by (b) IMS Nanofabrication.



(a) (b)

Figure 14. Schematic principles of the multi-beam columns from the tools proposed by (a) MAPPER and (b) IMS Nanofabrication

Other alternatives using multiple columns with one or several beams per column have also been proposed. Vistec Multi Shaped Beam (MSB) proposed a structure where each column contains a tool able to generate a Shaped Beam [STOLBERG, 2010]. Figure 15 shows an illustration of such structure.



(a) (b)

Figure 15. Schematic of (a) single variable shaped beam and (b) multi shaped beam technique, from [GRAMSS, 2010]

Another approach is employed by KLA-Tencor with its tool REBL [PETRIC, 2009a], [PETRIC, 2009b]. The main difference of this approach is that it is based on reflection. A large electron beam is projected over an electrostatic mirror matrix, each mirror being responsible for reflecting one pixel of the target image. Each mirror works like a blanker controlling the behavior of a single beam. The exposure is performed at several wafers at the same time by either a rotating stage (as shown in Figure 16 (a)) or a linear stage (as shown in Figure 16 (b)).

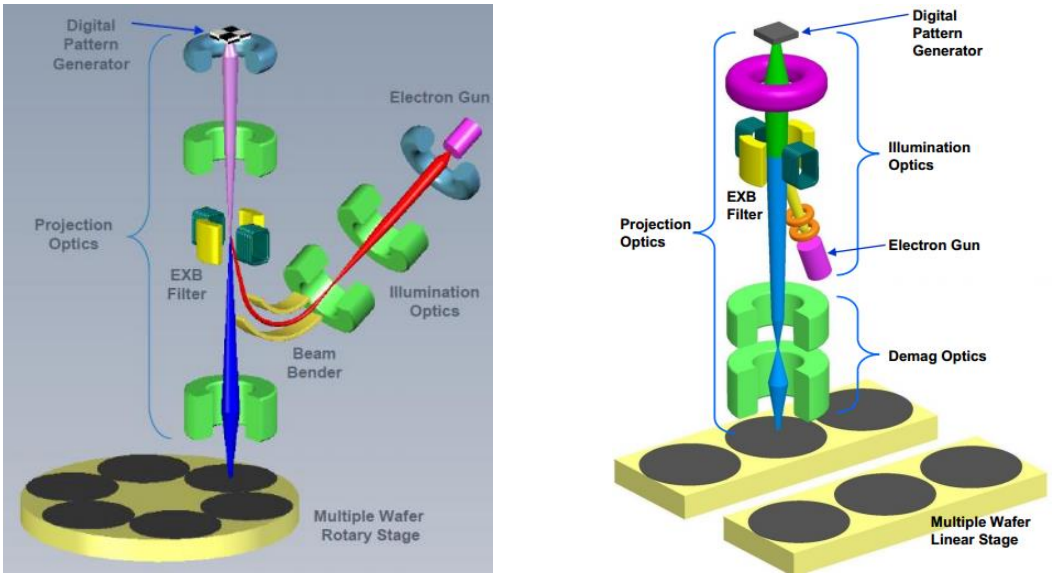


Figure 16. Schematic representation of the reflective tool REBL from KLA-Tencor

In comparison to the writing strategies of Gaussian-beam and Shaped-beam, the multi-beam approach writes several pixels (as in multiple Gaussian-beam) or several shots (as in multiple shaped-beam) at once. Figure 17 shows a graphical representation of the sequence of exposure of a multiple Gaussian-beam tool. Refer to Figure 12 (b) to compare with the standard single Gaussian-beam approach.

In any case, the multi-beam approach is still dealing with some important issues in order to be used by the industry. Notably, the exposed surface is heated due to the large number of beams reaching the same area at the same time and the beam placement is difficult to achieve accurately to control the dose received at each point. Moreover, another critical issue is the handling of such large volume of data due to pixelating the intended patterns, requiring a large data bandwidth.

Another approach considered by the industry to provide acceptable throughput in the lithography process, especially for regular designs such as memories, is the nanoimprint lithography (NIL). It is described in the next section.

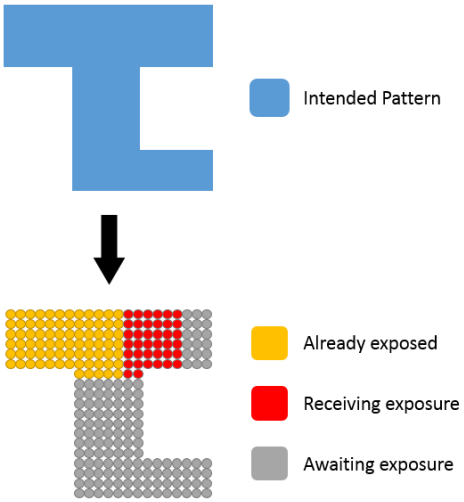


Figure 17. The intended pattern being exposed by a multi-beam exposure tool

1.2.5. Nanoimprint lithography (NIL)

Nanoimprint lithography (NIL) has emerged as a candidate solution for next-generation IC manufacturing for its advantages for mitigating the limitations of both photolithography and e-beam lithography. Thereby, nanoimprint comes with a promise of high-resolution as it does not suffer the diffraction issue that limits the photolithography and high-throughput in comparison to the e-beam lithography [CHOU, 1996][TORRES, 2003].

Basically, there are two main nanoimprint strategies: thermal imprint and UV-imprint. The first process consists in heating a thermoplastic polymer above its glass transition temperature, enabling the material flow to fill the structures of a mold. After, the temperature is lowered until the replicated patterns are solidified in place. Finally, the mold is removed. The UV-imprint employs an UV-curable monomer as compliant polymer layer in order to enhance fluidity of the printed material. The monomer layer is imprinted and submitted to broadband UV light radiation in order to form a rigid polymer. This enables the imprint pressure to be significantly reduced and avoid the time consuming and stress induced by a high temperature cycle. Figure 18 presents generic views of the (a) thermal nanoimprint and (b) UV-imprint lithography flows.

The mold is usually made of quartz or silicon, which is patterned by any conventional lithography technique. Usually e-beam lithography is preferable due to its higher resolution whenever the nanoimprint aims to achieve the requirements of advanced technology nodes.

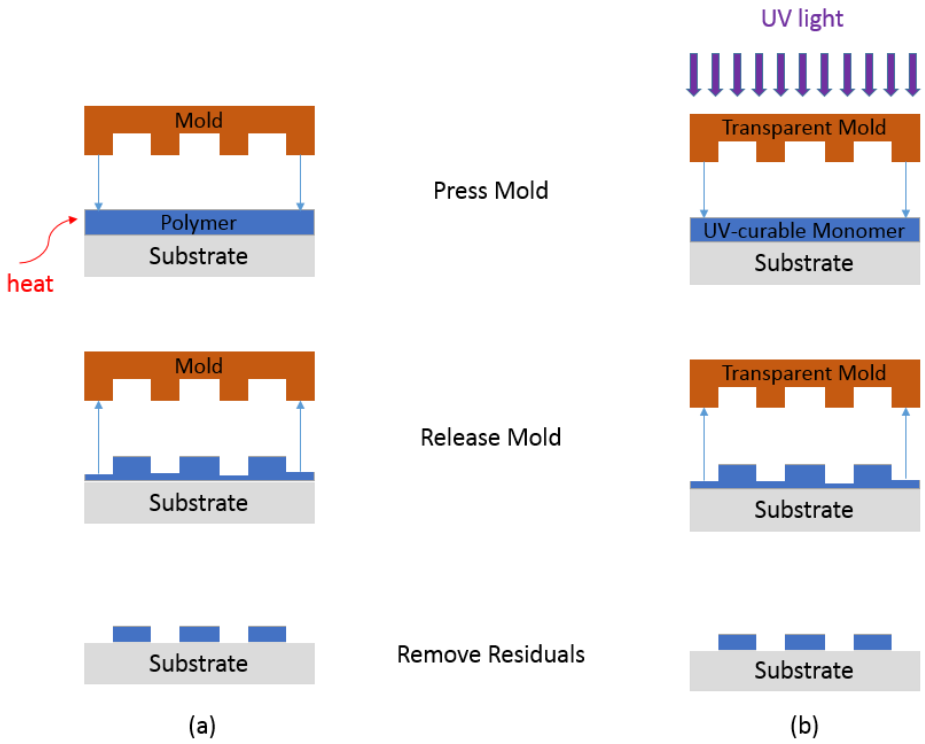


Figure 18. An illustration of the (a) thermal nanoimprint and (b) UV-imprint

Although nanoimprint is continually progressing, there are some issues that must be addressed. Adhesion between stamp and resist is an issue since the remains of the resist on the mold degrades pattern, reduces yield and damages the mold. Another critical issue is the life time of the mold. For the thermal nanoimprint, the changes of temperature and the high pressure cause stress and wear on molds and anti-sticking coating. This stress also presents a problem of alignment while using nanoimprint for multi-layer fabrication. Moreover, the viscosity of embossed material is an important issue, which may be a limiting factor for minimizing pattern dimensions and increasing pattern density. For UV-imprint, reactive species are generated during the light exposure. These species attack the anti-sticking layer of the mold so after a given number of imprints the mold must be cleaned to maintain sufficient demolding efficiency.

1.3. Lithography Process Comparison

It is not clear which technique will be selected to be the next lithography process. It is highly probable that different applications or different layers will use different strategies, as their benefits are sometimes complementary and the constraints are different. Multiple-patterning is the approach in use by the industry to continue to advance and a multiple-patterning EUV approach is already being discussed for the near future. Multi-beam machines are in different stages of development and are likely to increase in importance for writing the masks for the new strategies based in light and the molds of nanoimprint lithography (NIL). Nanoimprint is a strong candidate for very regular structures such as memory circuits while DSA is still in an early stage of development to evaluate how it will impact the future technology nodes lithography strategies.

Nonetheless, a clear aspect that is shared by all approaches is that they benefit from using e-beam exposure at some point on their flow, as summarized in Table 2.

Table 2. The different next generation lithography candidate approaches and where e-beam tools are used on their flows.

Strategy	Mask	E-beam tool used?
193i multiple patterning	Yes, more than one per complex layer	Yes, to the mask writing
Extreme-ultra-violet lithography	Yes, reflective with complex stack	Yes, to the mask writing
Directed self-assembly	Probably, for guiding structures and for cut-mask	Yes, to write the guiding structures and cut masks or for directly cutting the patterns
Multi-beam Direct Write	No, write the intended patterns directly on the wafer	Yes, to the wafer writing
Nanoimprint	No, use a mold to print the intended patterns on the wafer	Yes, for writing the mold

1.4. Conclusion

In this chapter it was shown that the current industrial technique for lithography (193 immersion optical lithography) is close to its limits and the industry is looking for a new approach. In order to mitigate the late availability (if ever) of EUV, the trend is to work with multiple patterning, which is already in production today for advanced technology nodes such as 18nm. However, this method presents its own limitations and is not able to provide, alone, a definitive answer for the technical and financial constraints faced by the industry for technology nodes below 14nm. Other strategies such as e-beam direct write using multi-beam technology and the directed self-assembly (DSA) approach are likely to be integrated to the flow for future technology nodes (either 10nm, 7nm or beyond). For some specific applications, such as memory chips, NIL is also a serious candidate to be used in future technology nodes. Nonetheless, it was shown that e-beam machines are the assembly point of all these alternatives, either as the main exposure tool or as the tool used to write the masks for DSA or optical lithography.

As e-beam machines seems to be down the road in any of the future strategies and even as the work horse for mask writing in current technology for the critical layers, it is imperative to understand the compensate for its limitations in order to achieve better resolution. Moreover, other factors that also limit resolution, such as etching bias and acid diffusion, must be addressed as well. The first step to compensate for these phenomena is the ability to understand, describe and predict the results of the lithography process. As these phenomena are more and more complex, the best manner to predict their behavior is to employ modeling techniques. On the next chapter the general concepts involved in modeling are discussed while further in this work specifics on lithography modeling are also presented and discussed.

Chapter II. Modeling

"All models are wrong, but some are useful"

- *George Box*

2. Modeling

A model may be seen as a representation of part of the physical world. In other words, it is a way to quantify, visualize or simulate a phenomenon. A model may be used to explain a system and to study the importance and the impact of different contributing factors. Moreover, it may also be used to make inferences of hypothetical conditions of the system it describes, i.e. prediction.

Mathematical models describe reality by means of mathematical functions. They may represent empirical relationships (compact models), physical relationships (first principle models) or a combination of empirical and physical relationships. Empirical relations are obtained by mathematical equations emulating observations while the physical relations come from the physical laws that apply to the system.

The study of a mathematical model may be divided in three parts: one that is related to aspects that are relevant to explain the system it describes i.e. its parameters or input variables; a second one which is how these parameters interact to impact the system, i.e., the equations; and a third one which is the result of these interactions, i.e. the output of the model. A simplified representation of a generic model could be Y = *f*(X) where Y is a vector of *m* outputs described as Y = [y_1 .. y_m] given by the computation of the equation *f* applied to the vector of *i* inputs X = [x_1 .. x_i].

Modeling is the process of producing, verifying, validating, using and comparing models. The diagram presented in Figure 19, first presented in 1979 by the Society for Computer Simulation, shows the primary phases and activities of computer modeling [SCHLESINGER, 1979].

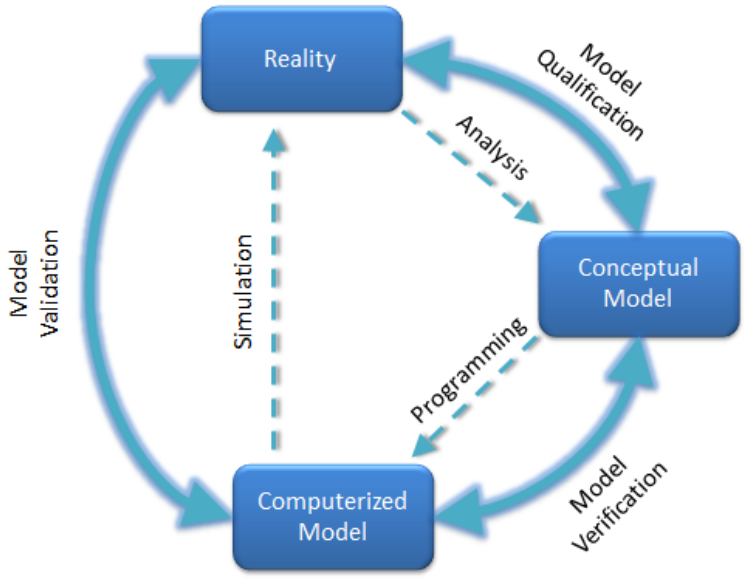


Figure 19. The primary phases and activities of computer modeling and their relations.

The diagram presents the phases and activities related to the generation of a model. However, whenever a model is being created based on experimental data, as oppose to ab initio models, the first step is to analyze this data in order to use only valid one for the model generation. This is discussed in section 2.1.

The model analysis process is used to conceive a conceptual/mathematical model of the reality. Programming converts the conceptual model into its implementation while simulation is used to estimate or predict reality. Moreover, the diagram presented in Figure 3 also presents the activities of model qualification, model verification and model validation, which are the drivers for the construction of a good model. Those activities are presented in sections 2.2, 2.3 and 2.4, respectively.

In addition, a good modeling also consists in choosing the most adequate model among several candidates. Model selection is further discussed in section 2.5.

2.1. Data Analysis

Before using experimental data in a modeling flow for model qualification, it is necessary to ensure that the data is consistent. In this sense, the main cause of data inconsistency is the existence of outliers. An outlier is an observation point that is distant from others due to experimental or measurement errors, for instance. Before discussing outliers and the ways to identify them, some definitions must be presented:

2.1.1. Quantiles and Quartiles

In order to analyze a set of data it may be interesting to go beyond the mean and variance. One strategy consists in analyzing the data set in portions. One way to divide a data set that is frequently used are the quantiles. Quantiles are values that divide a given distribution into regular steps in terms of number of elements. A quantile is not a partial data set but the exact value that stands dividing the distribution, either with an exact value of an element or as the mean between two elements. For instance, consider a data set of 12 elements (hence, an even sized set) which is to be divided into 3 equal parts by two quantiles, q1 and q2. Each quantile will then divide the set into two subsets, one containing 1/3 of the elements and the other containing 2/3. As the set has an even size, the quantile values will be the mean of the two closest values to the division point. This example is represented in Figure 20(a). When the data set size is odd, as represented in Figure 20(b), the quantile value is exact and it divides the data set in 1/3 and 2/3 excluding the point itself.

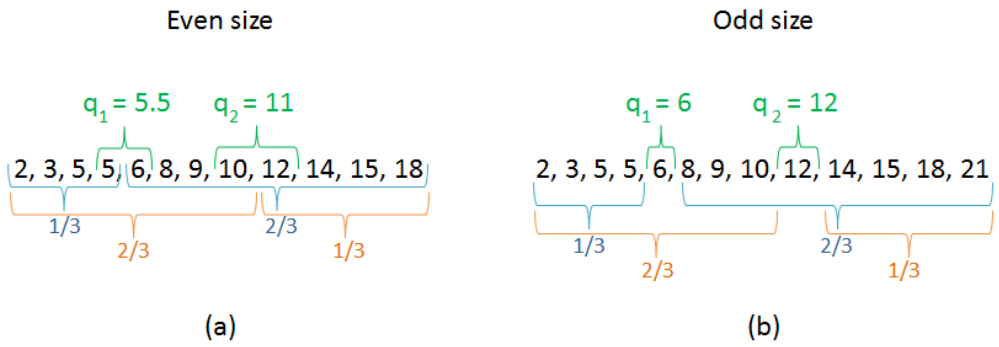


Figure 20. A representation of (a) an even size set of points and (b) an odd size set of points. In green, the two quantiles that divide the set into three equal sized sets. In blue is the division represented by the first quantile while in orange is the division represented by the second quantile.

Among all possible ways to divide a data set in quantiles, the more frequently used is to divide in four parts. This division is known as quartiles. The first quartile (Q_1), also called lower quartile, divides the set of data into the first 25% and the other 75%. The second quartile (Q_2), which is the median of the distribution, divides the data into two parts containing 50% of the data each while the third quartile (Q_3), also called upper quartile, divides the set of data into the first 75% and the resting 25%. A representation of a distribution and its quartiles is presented in Figure 21.

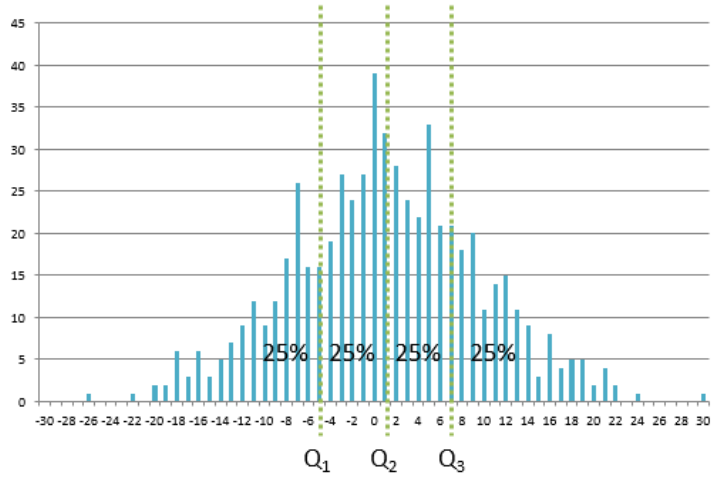


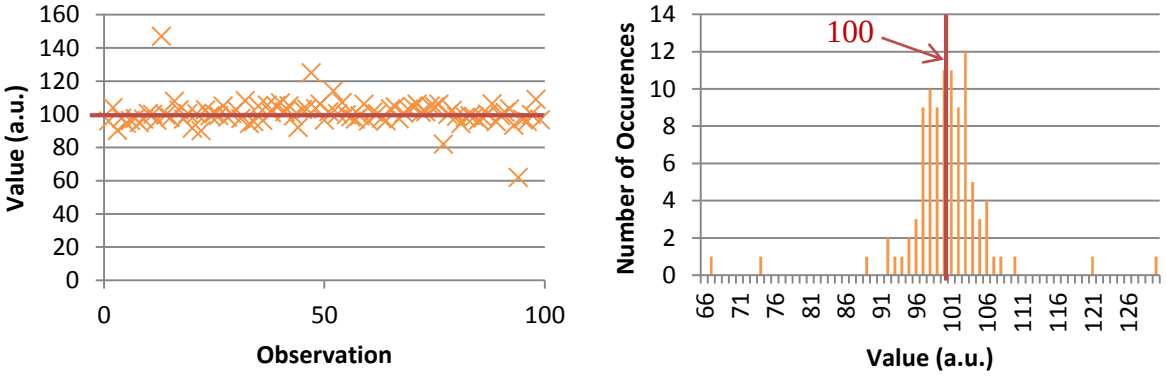
Figure 21. An example distribution and its first, second and third quartiles.

2.1.2. Univariate and Multivariate Distributions

The characteristics of a distribution play an important role in the type of strategy employed to detect outliers. One of the most relevant characteristics is if the distribution is univariate or multivariate. A distribution is considered to be univariate if it represents the probability distribution of a single variable. In other words, the data values behave as a single group, around a given mean value and presenting a certain variance according to any given law.

In order to illustrate the univariate distributions, consider the following example: one may determine how long it takes to boil water in a regular oven. The experience is repeated one

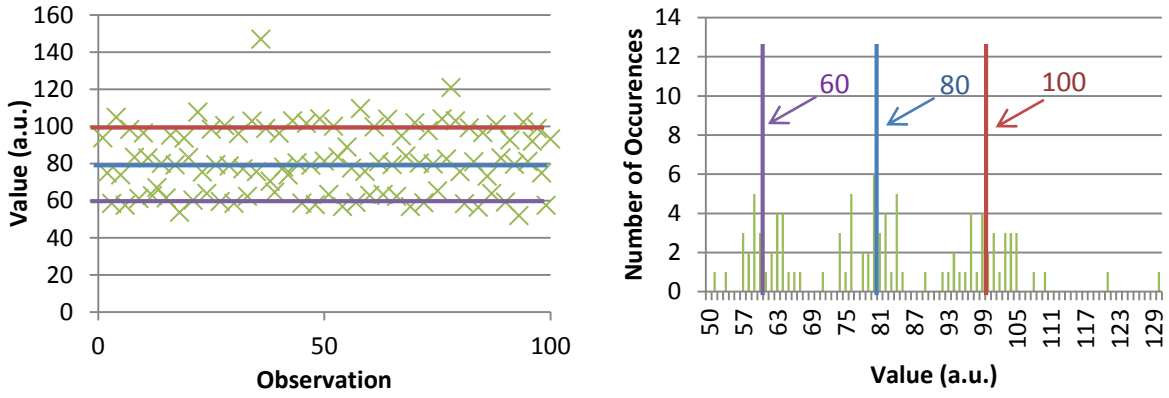
hundred times. This provides one hundred observations of the same phenomenon. The data set results presented in Figure 22 (a). From this set of observations, one may see that the expected experiment result is around 100, which is its mean value. All values are well centered to a mean value and the distribution seems to respect a single normal distribution, as it is possible to remark from its histogram, presented in Figure 22 (b).



(a) (b)

Figure 22. (a) An example of 100 observations of a given experiment and (b) the histogram of the same distribution.

Whenever a distribution represents the probability distribution of two or more variables, it is considered multivariate. An example of multivariate distribution is shown in Figure 23(a) while its histogram is presented in Figure 23(b). One may consider that the previous experiment was performed in three different ovens. By observing the data, one see that there is not a single particular value which all points are distributed around it, but three of them, 60, 80 and 100. This is even more clearly visualized in the histogram where the distribution seems to be composed of three peaks.



(a) (b)

Figure 23. (a) An example of 100 observations of a given experiment performed in different conditions and (b) the histogram of the same distribution.

2.1.3. Outliers

In data analysis, one of the most important source of errors is false data. It may be a result of an error in the experiment or just a random effect but, in either case, it will only disturb the

data analysis and lead to problems on the way a given phenomenon is understood or explained. For this reason, determining if a point should not be considered as part of a data set is important to assure data quality. Data points that differ from most of the observations are called outliers.

An outlier may be related to variability or error in the experiment or measurement. For instance, observe the distribution presented in Figure 22(a). It is possible to identify some points that are more distant to the mean than most of the others. These observations are strong candidates to be considered outliers.

The task of determining if a given observation is an outlier or not is complex and often relies on a priori knowledge of the data distribution or on the analysis of the experimental conditions that resulted in those observations. In some cases, where the size of the experimental data is large, it is prohibitive to analyze manually each data point in order to certify if it is valid or not. In the literature there are several automatic approaches to detect outlier values in a distribution [GRUBBS, 1969]. Most of those strategies are pruned either for univariate or multivariate distributions.

2.1.3.1. Outlier Detection for Univariate Distributions

The detection of outliers in univariate distributions are often performed based on evaluate the behavior of a statistic criterion. Some strategies are listed in [ULLRICH, 2010] and may be based on:

- Chi-squared score of data;
- Score of extreme value;
- Ratio of range to standard deviation;
- Ratio of variances with and without suspicious values;
- Ratio of ranges and subranges.

However, these analyses rely on the sample mean and/or standard deviation and assume that the distribution of the data without outliers respect a normal law.

There are also strategies that are based on robust statistics [MÜLLER, 2000]. In this case, instead of relying on mean and/or standard deviation, they are related to the median and calculate ranges without an a priori distribution law. One example is the use of the median absolute deviation (MAD) and the median value in the method known as MAD_e. MAD is a robust measurement of the variability of a univariate sample. It is defined as the median of the absolute deviations from the data median, calculated as presented in (4).

$$\text{MAD} = \text{median}\ (|val_i - median(val)|) \tag{4}$$

For instance, consider a set of values: S = [0, 2, 3, 5, 8]. Its median value is 3. The absolute differences from the median gives the set [3, 1, 0, 2, 5], which we may sort to [0, 1, 2, 3, 5] and identify the median of this set as 2. Therefore, the MAD value of the set S is 2.

In order to use this value for identifying and removing outliers, one may consider valid only the data inside the range given by (5) [SEO, 2006]:

$$[Q_2 - \omega\ \text{MAD}_e,\ Q_2 + \omega\ \text{MAD}_e] \tag{5}$$

Where Q_2 is the median value (second quartile) of the distribution and ω is a weighting factor depending on the aspired confidence level. The value of MAD_e is calculated as shown in (6).

$$\mathrm{MAD}_e = \kappa\,\mathrm{MAD} \tag{6}$$

Where κ is a constant scale factor used to estimate the standard deviation of a distribution from its MAD value. For instance, a normal distribution has its κ factor equal to around 1.4286, as demonstrated by Seo [SEO, 2006].

Another strategy, which is recommended by the NIST [NIST, 2014], is to use the Interquartile Range (IQR). The interquartile range (IQR), also called midspread, is a metric of dispersion and represents the difference between the upper and lower quartiles. IQR is calculated as presented in (7):

$$\mathrm{IQR} = Q_3 - Q_1 \tag{7}$$

An example of IQR is shown in Figure 24 based on the same distribution example presented in Figure 21.

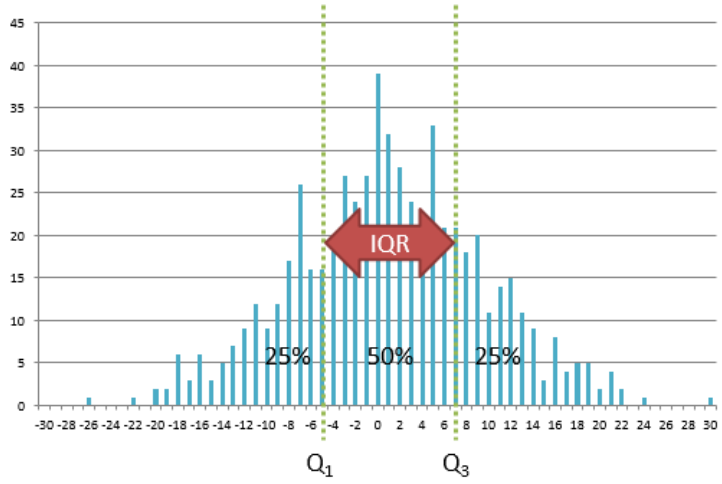


Figure 24. Same distribution from Figure 21 showing the IQR.

The method first described by Tukey [TUKEY, 1977] describes two types of outliers: extreme outlier is the one outside the range described in (8) for a ω equal to 3 while a mild outlier is the one inside the extreme range but outside the range for ω equal to 1.5.

$$[\mathrm{Q}_1 - \omega\mathrm{IQR},\ \mathrm{Q}_3 + \omega\mathrm{IQR}] \tag{8}$$

Where Q1 is the first quartile, Q3 is the third quartile, of the distribution and ω is a weighting factor depending on the aspired confidence level.

While the MAD strategy takes into consideration the entire data set, the IQR based strategy only provides information regarding the half of the elements in the distribution. This provides the IQR strategy with less sensitivity to outliers, since those values are often not included in the index calculation. On the other hand, considering only 50% of the data set to calculate the IQR value may be too restrictive for heavy tailed distributions.

2.1.3.2. Outlier Detection for Multivariate Distributions

Determining if a given observation is an outlier in a multivariate distribution is more complex than for univariate distributions. One approach is to separate this distribution into several univariate ones. This is feasible if the origin of this multivariation is known. Once the distribution is separated into several univariate ones, any of the techniques described in 1.2 may be used.

When there is no assumption of determining from which group an observation comes from, it is possible to estimate it. One technique for doing that based on fuzzy logic is presented in [TOP, 2011]. After evaluating by fuzzy logic the probability of each point to belong to each of the groups, they points are separated to create independent distributions. Each distribution is therefore univariate and the outliers may be removed based on any of the criteria presented on the previous section.

2.1.4. Data Variation (Noise)

Any experiment may be subject to variation. For instance, when an experiment is repeated in what is expected to be the exact same conditions, the results should be identical, which is rarely the case. Whenever the difference is punctual and very far apart from other observations, it is an outlier and may be handled as such. However, there is also variations that may affect every single observation and it is usually related to phenomena that cannot be fully controlled, such as material homogeneity or imprecision in the instruments used on the test and measurements.

Independently of the source of data variation, it may be useful to know its characteristics. One of the main interests of having information about the noise is to know the limits of precision one may expect from any model implemented to predict future realizations of the experiment. Moreover, it is interesting to know that any model that explains a data set with a deviation inferior to the data variation level is probably overfitting. Overfitting is explained in section 2.2.2.

Sometimes the variability of a given experiment may be estimated based on theoretical information from physical models and from equipment uncertainty information. Another approach is to repeat the experiment a certain number of times and observe the distribution of the observations.

2.2. Model Qualification

Mathematical models are usually composed of relationships and variables. Relationships can be described by functions, algebraic operators, differential operators, etc. Variables are abstractions of system parameters of interest, which can be quantified.

The first step for building a model is to identify the processes that have influence over the phenomenon. This step is usually performed by using experimental data, simulations of previous models or relying on experts. This step will provide which are the variables to be used and how they interact among themselves, i.e. the relationships of the model.

The second step is to determine the values of the model parameters. Usually this is performed based on experimental data. The transformation from the experimental data to model parameters may be performed by solving an inverse problem.

For some types of models, the inverse problem may be solved analytically. However, there are many types of models for which there is no direct way to obtain the parameter values. In these cases, one approach is to employ a parametric optimization method. This consists in finding the set of parameters minimizing a cost function, which describes the discrepancy between experimental data and simulation data from the model.

2.2.1. Cost function

A cost function assesses the difference between experimental data and model results. Its choice is a delicate aspect of the model calibration strategy as using different cost functions can lead to a different set of parameters. Some examples of cost functions are given in (9), (10) and (11).

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|exp_i - sim_i| \quad (9)$$

Mean Square Error

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(exp_i - sim_i)^2 \quad (10)$$

Mean Square Log Error

$$MSLE = \frac{1}{n}\sum_{i=0}^{n}(\log(exp_i) - \log(sim_i))^2 \quad (11)$$

Where n is the number of elements, exp_i is the i^{th} experimental data value and sim_i is the i^{th} model simulation result.

While the MAE present the same unit as the experimental data or the model result, both MSE and MSLE present this unit squared. For the sake of convenience, those values are sometimes used after extracting their square root and are respectively called RMSE (for Root Mean Square Error) and RMSLE (for Root Mean Square Log Error).

Several methods of cost function minimization exist, such as Levenberg-Marquardt [LEVENBERG, 1944], [MARQUARDT, 1963], the Simplex algorithm [NELDER, 1965], among others. However, the problem of most of these algorithms is that the search is performed locally. Therefore, the model parameter values obtained may not correspond to the global optimum. This is particularly true when the number of parameters to be optimized increases. One way to overcome this limitation is to employ global optimization methods such as the Efficient Global Optimization (EGO) [JONES 1998], genetic algorithms or simulated annealing [KIRKPATRICK, 1983].

2.2.2. Overfitting

Another important aspect in model qualification is to prevent a model from overfitting the data. Overfit occurs when a model describes random variability present in the data it is using for

calibration. It is usually related to the criterion for calibration that attempts to minimize the distance between observation data and the model predictions, which is represented by the cost function.

The risk of overfitting increases as the number of parameters of the model increases, as the number of observation in the data set used for model optimization decreases and as the model flexibility increases.

An example of data containing two different groups (triangles and circles) is given in Figure 25(a). Figure 25(b) shows a simple model calibrated to separate this two groups, leaving on the left the triangles and on the right the circles. Results show that this model fails to classify four elements (indicated in red). A second model is proposed, which is more complex, and is able to detect all points properly, as shown in Figure 25(c). Observe that the first model explains the general trend of the data with some errors while the second one gives exact result over the set. Therefore, if only the cost function is used to determine which one is the best model, the complex one would be selected. However, if a second sample if evaluated, such as the one shown in Figure 26(a), by used these two models, the results may be significantly different. Figure 26(b) shows the result from the classification of triangles and circles using the first model while Figure 26(c) shows the result for the second one. Observe that the first model continues to explain the general behavior with some errors while the second one generates much more false detections.

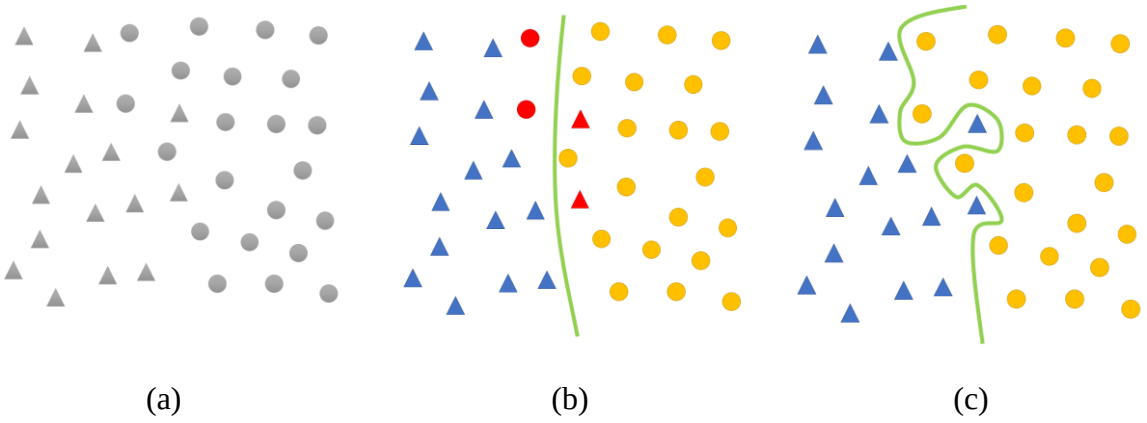
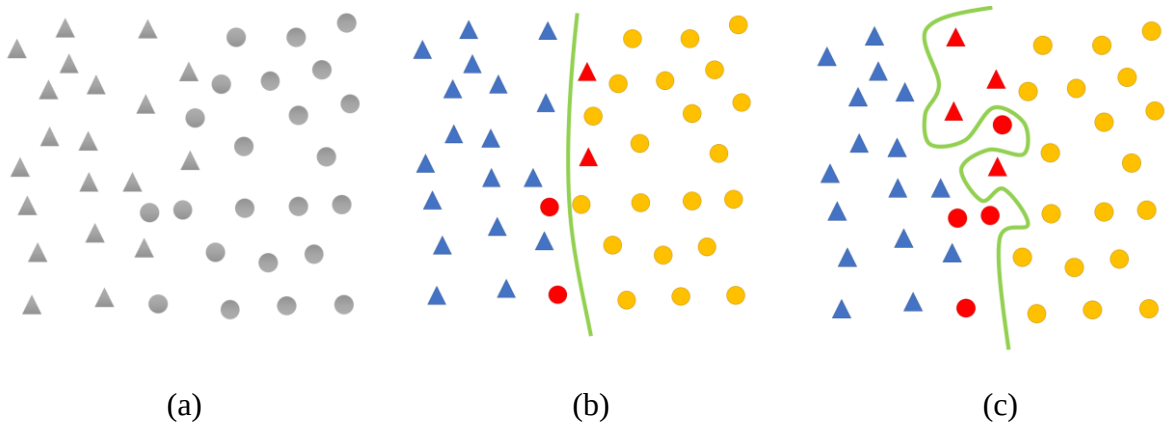


Figure 25. (a) Example of a distribution of points belonging to two different groups (triangles and circles). (b) Result of the detection based on a simple model (in green) tagging triangles in blue and circles in yellow. (c) Result of the detection based on an over-complex model (in green) Wrong detections are shown in red.

This observation also applies to whenever it comes to extrapolate the data set available. A model is unlikely to present good prediction to values outside the range of the observations if it is impacted by the noise present in the data. However, a simpler model that is not influenced by the random variability may be extrapolated and is expected to be more predictive.



(a) (b) (c)

Figure 26. (a) Example of a second distribution of points belonging to two different groups (triangles and circles). (b) Result of the detection based on the simple model (in green) calibrated in previous example. (c) Result of the detection based on the over-complex model (in green) calibrated in previous example. Wrong detections are shown in red.

In order to avoid overfitting, it is necessary to use additional techniques, either by model validation (presented in section 2.4) or by techniques of model selection (shown is section 2.5).

2.3. Model Verification

Verification is concerned with implementing the model right. The verification approach consists in evaluating if the results given by the implementation of the model are consistent with what is expected. To summarize, verifying a model consists in answering the question: Is the model implemented correctly?

A model may be said verified whenever the following assertions are true:

- The model is programmed correctly;
- The algorithm have been implemented correctly;
- The model does not contain errors, oversights or bugs.

Therefore, verification ensures that the specification is complete and that mistakes have not been made during the implementation of the model. However, it is important to remark that the verification process does not ensure that the model correctly reflects the system it represents. This is the role of model validation.

Since it is not practical to fully verify a model, the goal must be to present a high degree of statistical certainty. This may be assured by testing important cases, or all possible cases, and by employing automated testing processes.

2.4. Model Validation

Validation is concerned with building the "right" model. The validation approach consists in evaluating if the results given by the implementation of the model are an accurate enough representation of the system. The ultimate goal of validation is to make the model useful in the sense that it addresses the right problem, providing the proper information about the system.

As an aid in the validation process, Naylor and Finger [NAYLOR, 1967] formulated a three-step approach which has been widely followed:

Step 1 - Build a model that has high face validity: Face Validity is the process of identifying a model as coherent and adequate to explain and predict a given system. Generally, this is performed by experts on the domain where the model is applied. If a model succeeds in a Face Validity test it is said that it "looks like" it will work. For instance, consider a model for identifying the risk of a driver to be involved in a car accident. If the model indicates that if the car speed goes up the risk of accident goes down, this model should be looked with care and may not present high face validity, unless experts can identify other factors that may justify this behavior.

Step 2 - Validate model assumptions: The assumptions of a model may be divided into data assumptions and structural assumptions. The first one is based on data collection and statistical analysis while the second one involves how the system operates including simplifications and abstractions of the reality. For instance, a data assumption consist in determining that the amount of water accumulated in a road is never superior to a given threshold while a structural assumption is to state that the water level on the road is not relevant to the system being modeled.

During the process of validation it is important to certify that the assumptions are true for the phenomenon being modeled. Otherwise, the system being described is outside the domain of validity of the model. For instance, for the previous model of the risk of car accident, consider that the presence of snow was neglected. If this model is to be used in north of Brazil this is a valid assumption that won't affect the results. However, if this model is to be used in the Alps region in France it will probably not be a valid assumption.

Step 3 - Compare the model input-output transformations to corresponding input-output transformations for the real system. The validation of the input-output transformations of the model in relation to the real system is basically the testing of the ability to predict future behavior of the real system. This step is performed by using real system parameters and comparing the model output result with what actually happened in the real system in the same situation. This evaluation may be performed using a cost function, as described in the previous section.

In order to perform the third step of this validation process, several strategies were proposed in the literature. The following subsections briefly describe the most common ones:

2.4.1. Cross-validations

The Cross-validation technique is based on the principle of using a data set for validating a model which was not used during the model calibration [KOHAVI, 1995]. This process consists of several rounds. On each round, all data available is divided in two sets, one to be used in the calibration procedure, called training set, and another to be used in the validation procedure, the testing set. This procedure is repeated several times, and at each time the data is divided differently.

There are several different strategies of cross-validation, which are all variations of the generic case called *k-fold cross-validation*. In this approach, the original sample is randomly divided into k equal subsamples. One subsample is designated as the testing set and the other k-1 subsamples are used as the training set. The cross-validation process is then performed k times, and at each time a different subsample is used as the testing set.

Another cross-validation technique is called *repeated random sub-sampling* validation [GEISSER, 1993]. In this approach the dataset is randomly split into training set and testing set. The procedure may be repeated as many times as desired and the result is then averaged over the splits. The advantage of this method over the k-fold cross validation is the proportion of the training/testing sets is independent of the number of iterations to be performed. However, using this method may imply that some observations may never be used for validation while others may be selected several times. Although this may cause no impact, in the case of the presence of some outliers on the data it is useful to be sure that these points were not always used on the calibration set.

2.4.2. Holdout validation

Holdout validation is another validation technique that consists in choosing randomly from the original dataset a subset to be used as the testing set, while the rest is used as the training set [DEVROYE, 1979]. This strategy is not considered as cross-validation since each observation is used either for training or for testing and there is no iteration involved.

2.4.3. Model Validation Summary

The choice of the strategy of model validation is mostly dependent of the amount of data available, the cost to obtain more data than necessary for the model qualification, and the time necessary for optimizing the parameters of the model. In summary, the decision could follow the given rules:

1. Cross-validation is an option when the optimization cost is not so expensive as to make multiple optimizations prohibitive. The number of rounds used and the decision between the k-fold strategy or the random sub-sampling depends on the amount of time available and the amount of time necessary for performing each round;

2. Holdout validation is an option when the data available is so abundant that keeping part of it out of the optimization is not an issue. During the experiment design it is possible to provide intermediate cases for validating interpolation behavior and extreme cases to validate the model behavior in case of extrapolation;

2.5. Model Selection

Model selection is the ability to compare different models and decide which one is more suitable for performing estimation or prediction of a given problem. According to Occam's Razor principle, among models with similar predictive power, the simplest one is the most desirable one.

This section describes a set of metrics available to compare different models based on the statistical data of the errors distribution. These criteria may be divided into two major categories, one composed of statistical estimators and another composed of robust estimators.

2.5.1. Statistical Estimators

2.5.1.1. Cost Functions

The first metric that may be used to compare models is the resulting value given by the cost function employed during the optimization procedure. Examples of cost functions were given in section 2.2.1. This criterion will select the model with the smaller difference between prediction and data. However, this choice may benefit very complex models, which may overfit the data used for calibration. This problem could be solved by applying the cost function over a second set of data, not used during model calibration. Nevertheless, the cost functions also do not take into account the differences between the model structures and complexities, i.e., the number of parameters used to explain the data.

2.5.1.2. Variance (σ2) and Standard Deviation (σ)

The Variance (σ^2) of the output error of a model may be used to compare its quality in relation to other models. It measures how far a set of numbers is spread out in relation to its mean. In other words, a model whose output errors present a small variance is more likely to be good in comparison to a model with a large spread. The variance is calculated as presented in (12):

$$\sigma^2 = \frac{1}{n}\sum_{i=0}^{n}(val_i - mean)^2 \tag{12}$$

The Standard Deviation (σ) of the output error is the root square of the Variance. The advantage of working with the standard deviation instead of using the variance is the units; while the first has the same unit as the error metric the second one present this unit in squares which, sometimes, make it less meaningful for further analysis for some models.

However, neither the Variance nor the Standard Deviation provides information about the mean error. Selecting a model based only on the Variance may lead to preferring a model that properly represents the behavior of the experimental data but has a constant error. Therefore, the Variance or the Standard Deviation are usually used in association with a cost function which provides the mean error information.

2.5.1.3. Minimum and Maximum Errors

One of the simplest ways to evaluate the quality of a model, or to compare two or more models among themselves, is to verify the extreme positive and negative prediction errors a model produces in regards to experimental results.

The problem with this evaluation is that it may be biased by the presence of outliers in the experimental data. In this case, this metric will benefit a model that better explains a false data and, therefore, with no relation to future occurrences and the prediction quality of the model. However, once the degree of confidence in the quality of the experimental data is high, it is important to understand the extreme prediction errors a model may produce since it directly affects the quality of a model.

2.5.1.4. Interquartile Range (IQR)

The interquartile range (IQR), also called midspread, is a metric of dispersion of the errors of a model estimation and the data. It is calculated as presented in 2.1.1 and represents the difference between the upper and lower quartiles. IQR is calculated as presented in (8).

Whenever evaluating the output of a model, the IQR of the error distribution may indicate the dispersion of the best 50% of the results. This may be not enough for evaluating the quality of the model, since it neglects the other 50% of the experimental data. Nonetheless, it may be used as a complement to the metrics that take into account the entire data set and, in this way, give a perception of the results sure to not be taking into account the outliers of the experimental data.

2.5.1.5. Minimum and Maximum Errors in Range

The interquartile range (IQR) may also be used to give a range were the minimum and maximum errors of a distribution may be searched. In this case, it is possible to overcome the drawback of the minimum and maximum errors strategy presented in section 2.5.1.3, which is the bias coming from the presence of outliers in the experimental data.

The strategy consists in searching the minimum and maximum errors excluding all points that fall below Q1 – 1.5(IQR) or above Q3 + 1.5(IQR). Figure 27 shows the same distribution presented in Figure 21 but including the Q1 – 1.5(IQR) and the Q3 +1.5(IQR) ranges and the minimum and maximum value within this range.

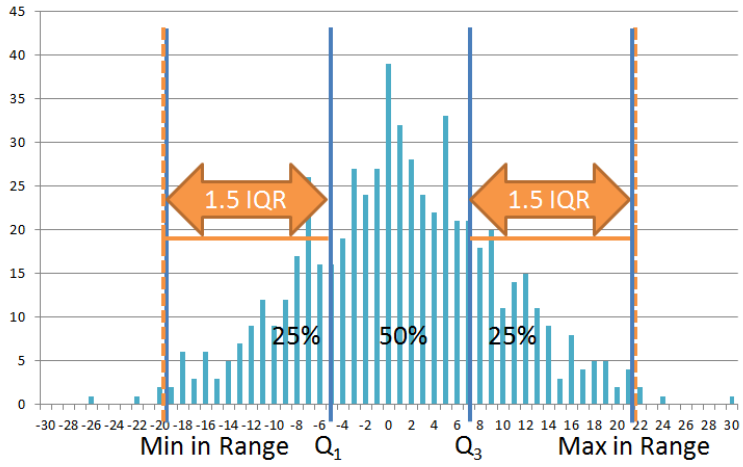


Figure 27. Same distribution from Figure 21 showing the minimum and maximum values inside the Q1 – 1.5(IQR), Q3 + 1.5(IQR) range.

Observe that by using this strategy the extreme points are disregarded. If the quality of the data is unsure, doing so is a way to prevent some outliers to bias the model evaluation.

2.5.2. Robust Estimators

2.5.2.1. Information Criteria

The evaluation of models using information criteria is founded on the information entropy. It offers an estimation of the loss of information when a given model is used to represent the experiment that generated the reference data. These strategies deal with the trade-off between the quality of the calibrated model and its complexity, given by the number of parameters used. However, these methods do not provide information about the absolute quality of a model but allows comparing among a set of candidate models.

Generally, any information criteria (IC) could be described as presented in (13).

$$IC = -2\,ln(L) + P.k, \tag{13}$$

Where k is the number of parameters, P is a penalty coefficient and L is the maximized value of the likelihood function.

A likelihood function is a function of the parameters of a model. In other words, it indicates how likely a given set of parameters is to be the true given a set of observations. In the case of a classical regression model with normally distributed errors, the first term $-2\,ln(L)$ may be replaced by $\mathrm{n}.\ln(MSE)$, where MSE is the mean square error and n is the number of observations, i.e., number of elements in the data set used to compute MSE. In this case, (13) becomes the equation presented in (14):

$$IC = \mathrm{n}\, ln(MSE) + P.k, \quad (14)$$

Where k is the number of parameters, P is a penalty coefficient, MSE is the mean square error and n is the number of elements in the data set.

Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is an evaluation of the quality of a model [AKAIKE, 1973]. AIC is calculated as the formula presented in (13) using a penalizing coefficient P equals to 2. In the case of the classical regression, the AIC formula is equal to the one presented in (15).

$$AIC = \mathrm{n} * \ln(MSE) + 2k, \quad (15)$$

Where k is the number of parameters, MSE is the mean square error and n is the number of elements in the data set.

Corrected AIC (AICc)

The Corrected AIC (AICc) is a version of AIC with a correction for the sample size and a greater penalty for extra parameters [HURVICH, 1989]. Whenever the number of samples n >> k, AICc becomes equivalent to AIC. As for the AIC, the smaller the AICc value, the better the model is compared to other models. AICc is calculated as shown in (16).

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}, \quad (16)$$

Where AIC is the Akaike Information Criterion calculated as shown in (15), k is the number of parameters and n is the number of samples being used.

Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) is a criterion for model selection closely related to the AIC. It was first presented by Schwarz in 1978 [SCHWARZ, 1978]. As in AIC, BIC also presents a penalty term for the number of parameters used in the model. Moreover, the BIC also includes a penalty regarding the number of samples used. Therefore, the BIC is calculated as the formula presented in (13) using a penalizing coefficient P equals to ln(n). In the case of the classical regression, the BIC formula is equal to the one presented in (17).

$$BIC = \mathrm{n} * \ln(MSE) + ln(n)\, k, \quad (17)$$

Where k is the number of parameters, MSE is the mean square error and n is the number of elements in the data set.

Hannan-Quinn Information Criterion (HQIC)

The Hannan-Quinn Information Criterion (HQIC) is another criterion for model selection. As BIC, it takes into account the number of samples used. However,

the number of parameters has a small effect compared to BIC or AIC. HQIC penalizing coefficient P equals 2 ln(ln(n)) and its general formula for the normally distributed errors hypothesis are given in (18).

$$HQIC = \mathrm{n} * \ln(MSE) + 2\,k\,ln(\,ln(n)), \tag{18}$$

Where MSE is the mean square error and n is the number of elements in the data.

Given a set of candidate models for the data, the preferred model is the one with the minimum IC value. However, this approach does not provide any information about the quality of the selected model. For instance, if the IC is used to select one out of a group of proposed models where none is able to properly represent the measured data, the IC method will nevertheless provide an answer indicating the one that better represents the data. For this reason, IC should never be used as the only selection criteria.

All candidate models are compared against the one presenting the minimum IC value. Generally, when comparing two models using a given selection criterion, models with similar scores should be placed in the same ranking category and that other criterion should be used to determine which of those models are better. Although there is no definitive rule on how to compare different IC values, some authors provided guidelines for AIC [BURNHAM, 2002] and BIC [KASS, 1995]. These guidelines are summarized in Table 3.

Table 3. Interpretation of the IC value differences between a given model *i* and the best model available presenting IC_{min} for AIC and BIC metric.

<table>
<tr><th>$IC_i - IC_{min}$</th><th>AIC</th><th>BIC</th></tr>
<tr><td>0 – 2</td><td>Models should not be set apart based on this criterion</td><td>Models should not be set apart based on this criterion</td></tr>
<tr><td>2 – 4</td><td rowspan="3">It is likely the Model min should be supported instead of Model i</td><td rowspan="2">There is some evidence against Model i</td></tr>
<tr><td>4 – 6</td></tr>
<tr><td>6 – 7</td><td rowspan="2">It is very likely that Model i may be disregarded</td></tr>
<tr><td>7 – 10</td><td></td></tr>
<tr><td>> 10</td><td>The Model i could be disregarded based on this difference</td><td>The Model i could be disregarded based on this difference</td></tr>
</table>

The way to read this table is the following. Any model presenting a AIC score difference in relation to the reference minimum one below or equal to 2 should be considered as good as the reference one. For differences between 4 and 7, the support for the model under comparison is not that high, and one could consider that the reference model is likely to be better than the one under analysis. Finally, if the difference is above 10, there is essentially no reason to consider the model under evaluation as a worthy candidate to replace or to continue to compete against the reference minimum one. Observe that this table presents some gray areas, like a difference of 3 or 8, where it is not clear what choice should be made. It is possible to

consider those cases as inconclusive and that other criteria should be used to either increase or decrease the support for the candidate model. When it comes to BIC, the table provides very similar information but, if in one hand it provides another intermediate degree of confidence, there are no gray regions.

2.5.2.2. Coefficient of Determination (R^2)

The coefficient of determination, also known as R^2, is a metric to indicate how well a set of data points is explained by a model output. In a general form, R^2 can be seen to be related to the ratio between the variance explained by the model and the one that is not explained by the model. In this case, the closest to one the value of R^2 is, the better the model explains the experimental data.

Consider a given experimental data set containing n samples, [exp_1 to exp_n], whose mean is given by $mean_{exp}$. Consider that a given model, trying to explain this experimental set, provides n simulation results, given by [sim_1 to sim_n]. For these two sets, three values are calculated as shown in (19), (20) and (21), respectively.

Total Sum of Squares: $SS_{total} = \sum_{i=1}^{n}(exp_i - mean_{exp})^2$ (19)

Explained Sum of Squares: $SS_{explained} = \sum_{i=1}^{n}(sim_i - mean_{exp})^2$ (20)

Residual Sum of Squares: $SS_{residual} = \sum_{i=1}^{n}(exp_i - sim_i)^2$ (21)

The R^2 coefficient is then calculated as the ratio between the Explained Sum of Squares and the Total Sum of Squares. Its equation is shown in (22).

$$R^2 = \frac{SS_{explained}}{SS_{total}} \quad (22)$$

As the total sum of squares is the sum of the explained and the residual ones (23), R^2 can be defined by its most common expression, given by (24).

$$SS_{total} = SS_{explained} + SS_{residual} \quad (23)$$

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} \quad (24)$$

In the best scenario the residual variance is zero and, therefore, R^2 equal to one. In the worst scenario the total variance is equal to the residual variance, making R^2 equal to zero.

2.5.2.3. Adjusted R^2

Although R^2 provides the information of how much of the experimental data variance is explained by the model, it does not take into account the number of parameters in the model being compared. Therefore, to prevent overfitting, Adjusted R^2 was proposed to take into account differences in model complexity [THEIL, 1961]. It is calculated as presented in (25):

$$Adjusted\ R^2 = 1 - \frac{n-1}{n-(k+1)}\ (1 - R^2), \tag{25}$$

This formula comes from the understanding that R^2 is a proportion between variances. Adjusted R^2 then replaces the classical description of variance in R^2 by their unbiased counterparts, as shown in (26) and (27).

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - mean)^2, \tag{26}$$

$$\sigma_{res}^2 = \frac{1}{n-1-k}\sum_{i=1}^{n}(m_i - x_i)^2 \tag{27}$$

Adjusted R^2 converges towards R^2 when the number of elements in the experimental data set is much larger than the number of parameters.

2.5.3. Correlation Matrix

Another relevant criterion when analysis a model is how the input parameters relate to each other. Ideally, all parameters should be independent. One strategy to evaluate the interaction among parameters is the correlation matrix.

A correlation matrix of the parameters of a model groups the correlation coefficient of each pair of parameters. The correlation coefficient indicates the influence that one parameter has over the other. For instance, a correlation matrix of a model presenting three parameters: x_1, x_2 and x_3, would be as shown in Table 4:

Table 4. Generic correlation matrix for a model with three parameters

	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$
$\boldsymbol{x_1}$	$Cor\ (x_1, x_1)$	$Cor\ (x_1, x_2)$	$Cor\ (x_1, x_3)$
$\boldsymbol{x_2}$	$Cor\ (x_2, x_1)$	$Cor\ (x_2, x_2)$	$Cor\ (x_2, x_3)$
$\boldsymbol{x_3}$	$Cor\ (x_3, x_1)$	$Cor\ (x_3, x_2)$	$Cor\ (x_3, x_3)$

Where $Cor\ (x_1, x_2)$ corresponds to the correlation coefficient indicating the influence of x_1 over x_2.

There are several ways to calculate the correlation among two variables [SPEARMAN, 1904]. One of the classical approaches is to use the Pearson product-moment correlation coefficient [PEARSON, 1895], which is calculated as presented in (28):

$$Cor\ (x_j, x_k) = \frac{Cov\ (x_j, x_k)}{\sqrt{\sigma_{x_j}^2 . \sigma_{x_k}^2}} \tag{28}$$

Where $\sigma_{x_j}^2$ is the Variance of the parameter x_j and $Cov\ (x_j, x_k)$ is calculated as presented in (29).

$$Cov\left(x_j, x_k\right) = \frac{1}{n}\sum_{i=0}^{n}\left(x_{j,i} - mean_j\right).\left(x_{k,i} - mean_k\right) \quad (29)$$

Therefore, the correlation coefficient value varies from -1 to 1. The module of the correlation coefficient indicates how the two parameters are correlated while the sign indicates if the correlation is direct or inverse. Some examples of values of two parameters and their correlation coefficient values are presented in Figure 28. Notice that for a correlation of 0.8 the relation is clear while for 0.4 the relation between x_i and x_j is less obvious.

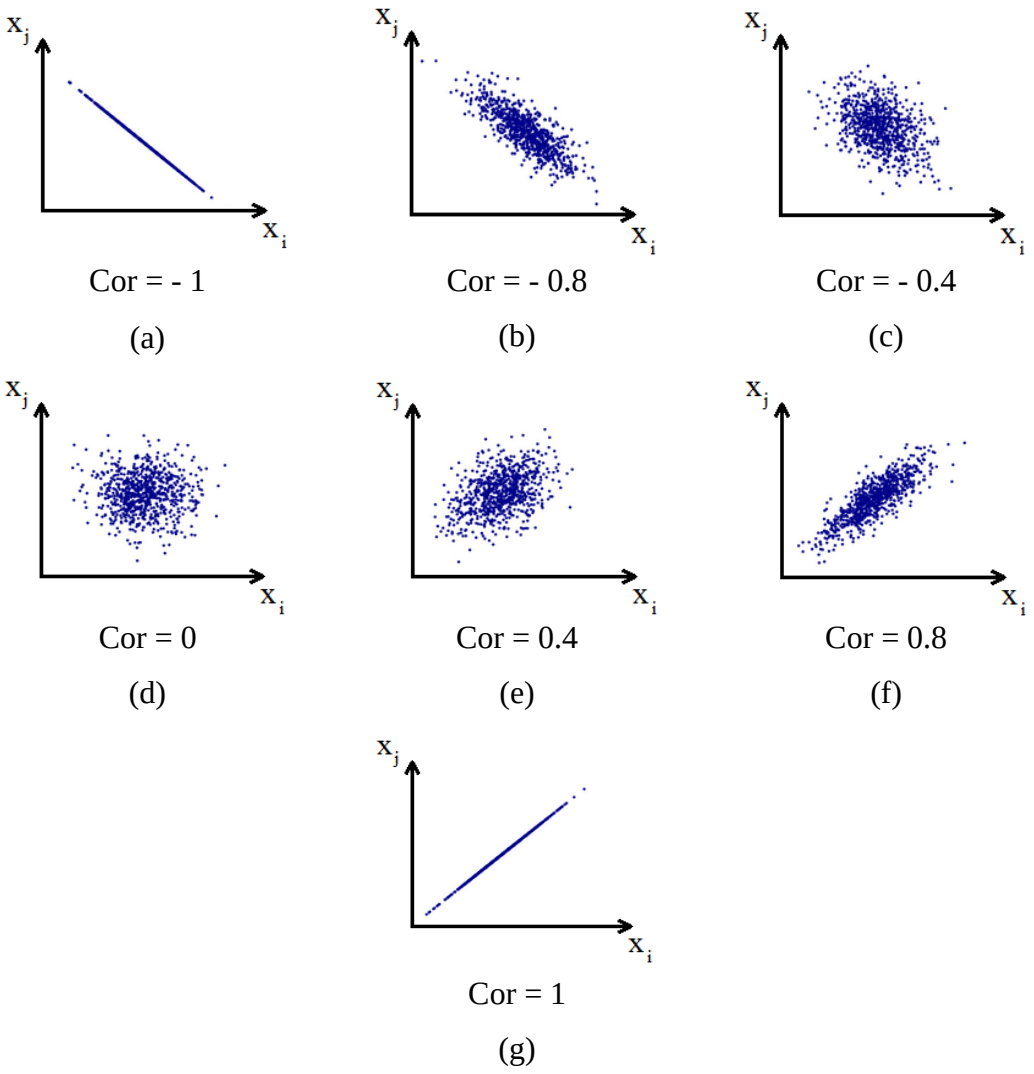


Figure 28. Seven different distributions of values for two parameters and their respective correlation coefficient value calculated as presented in (28).

The correlation coefficient represents how much correlation exists between two parameters but it does not provide information about the intensity of it. For instance, it does not matter if a small change in one parameter has an important impact over the other, if both parameters are linear correlated, the coefficient is going to be equal to 1, as represented in Figure 29.

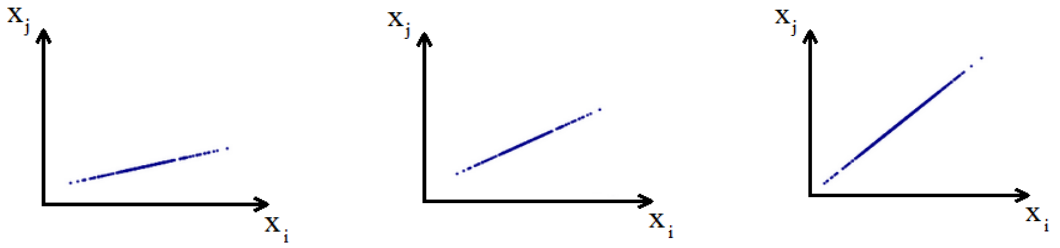


Figure 29. Three different distributions of values for two parameters and their respective correlation coefficient value calculated as presented in (28).

Nonetheless, it is important to remark that two parameters that are weakly correlated always present a correlation coefficient index close to zero but the inverse is not true. Strongly correlated variables may also present values of correlation coefficient close to zero, as long as their correlation is not linear. For instance, observe the distributions presented in Figure 30. In Figure 30 (a), there is no evident relation among x_i and x_j. In Figure 30 (b) and (c), they are related, but not in a linear manner and, therefore, the correlation coefficient calculated as shown in as in (28) are also equal to 0.

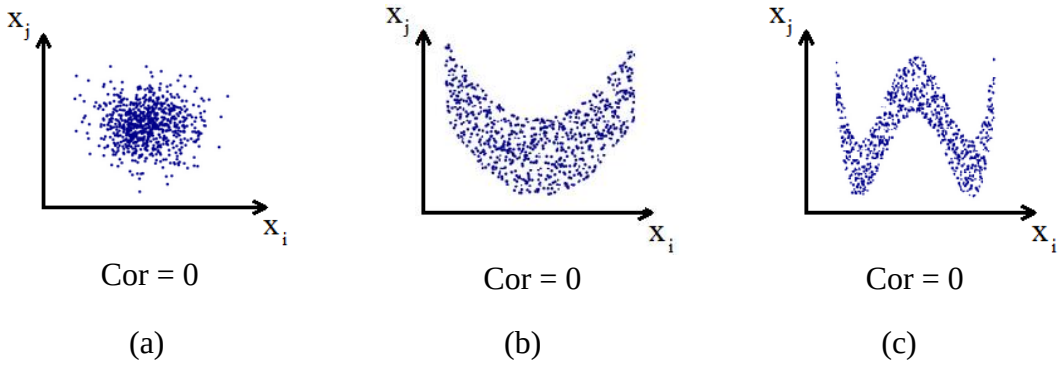


Figure 30. Three different distributions of values for two parameters and their respective correlation coefficient value calculated as presented in (28).

There are some strategies that may address correlation other than linear, such as the Spearman's rho correlation [SPEARMAN, 1904], which relaxes the linearity assumption to a monotonic one and the Brownian correlation [SZEKELY, 2007] which presents correlation as a distance and its value is only zero if the random vectors are independent.

Observe that from the equations (28) and (29) it is possible to derive the properties presented in Table 5.

Table 5. Properties of the Correlation Coefficient Pearson product-moment correlation coefficient

Property	Proof
$\boldsymbol{Cor\,(x_j, x_j) = 1}$	$Cor\,(x_j, x_j) = \frac{Cov\,(x_j, x_j)}{\sqrt{\sigma_{x_j}^2 . \sigma_{x_j}^2}} = \frac{\frac{1}{n}\sum_{i=0}^{n}(x_{j,i} - mean_j).(x_{j,i} - mean_j)}{\sigma_{x_j}^2}$ $= \frac{\sigma_{x_j}^2}{\sigma_{x_j}^2} = 1$
$\boldsymbol{Cor\,(x_j, x_k)}$ $\boldsymbol{= Cor\,(x_k, x_j)}$	$Cor\,(x_j, x_k) = \frac{Cov\,(x_j, x_{jk})}{\sqrt{\sigma_{x_j}^2 . \sigma_{x_k}^2}} = \frac{\frac{1}{n}\sum_{i=0}^{n}(x_{j,i} - mean_j).(x_{k,i} - mean_j)}{\sqrt{\sigma_{x_j}^2 . \sigma_{x_k}^2}}$ $= \frac{\frac{1}{n}\sum_{i=0}^{n}(x_{k,i} - mean_j).(x_{j,i} - mean_j)}{\sqrt{\sigma_{x_k}^2 . \sigma_{x_j}^2}} = Cor\,(x_k, x_j)$

Based on the properties from Table 5, the same correlation matrix presented in Table 4 may be presented simplified as shown in Table 6:

Table 6. Generic correlation matrix for a model with three parameters, simplified considering the properties presented in Table 5

	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$
$\boldsymbol{x_1}$	1	$Cor\,(x_1, x_2)$	$Cor\,(x_1, x_3)$
$\boldsymbol{x_2}$	-	1	$Cor\,(x_2, x_3)$
$\boldsymbol{x_3}$	-	-	1

The interest on evaluating the correlation among parameters after optimization is to certify that the result provided is robust. If two parameter values are strongly correlated, it means that both may be miscalculated and one error is being compensated by the other. It can be shown that the uncertainty on the determination of the parameter degrades when correlation increases.

2.5.4. Model Selection Synthesis

Regarding the statistical criteria:

Usually, when the degree of confidence in the quality of the data is high, the use of MSE or RMSE is sufficient to provide statistical information about the quality of the model. They can always be combined with the Minimum and Maximum Errors information to provide a clear idea of the extreme cases. However, if the quality of the data is questionable (either due to outliers or a high variability), one may prefer to use strategies that are less biased by the error values, such as the MAE, and then combine it with the Variance/Standard Deviation values for providing information about the spread and the Minimum and Maximum Errors in Range for the extreme cases.

Regarding the information criteria:

One of the most frequent comparison between criteria in the literature regards AIC and BIC [BURNHAM, 2002][BURNHAM, 2004][TU, 2012][YANG, 2005]. What is generally stated

is that BIC criterion penalizes more the models with a larger number of parameters in comparison to AIC. Moreover, authors tend to prefer AIC arguing that it is asymptotically optimal in selecting the model with the least MSE even if the exact model is not in the candidate set [YANG, 2005]. In this case, there is a net preference for AICc over other information criteria.

On what concerns HQIC, it presents the same characteristics of BIC in most cases, but it is less effective in preventing overfit due to its smaller penalty factor for small data sets. Therefore, HQIC is not much used in practice for model selection [BURNHAM, 2002].

In order to choose which information criteria to use when comparing several models, here follows some tips:

1. AICc can always be used instead of AIC. When the data set is small, AICc will compensate and help preventing overfit while AIC does not take directly the number of samples into account. For large data sets, AICc converges toward AIC.
2. The penalization regarding the size of the data set is always more significant for the BIC than for the HQIC.
3. Whenever the cost function is related to MSE, AICc should be used instead of the BIC.

Regarding the other mentioned methods:

The coefficient of determination, or its advanced variant, provides a piece of information that is not provided by any previously mentioned criteria, which is the relation of explained and unexplained variance of the experimental data. The advanced variant is preferable due to the penalty it includes for the amount of parameters used in relation to the size of the experimental data set. When the number of elements is large enough compared to the number of parameters they are equivalent, hence the advantage of always using the advanced variant.

Finally, the correlation matrix also provides a piece of information that lacks in all other criteria, which is the correlation between two model parameters. A model that has weak correlation among its parameters is usually preferable when compared to a model presenting strong correlations. This may not be a first order selection criterion, but may be taken into account whenever it comes to decide if a selected model is good enough or more models should be created/calibrated/considered for the provided experimental data.

2.6. Conclusion

The first step for building a model is to identify the processes that have influence over the phenomenon. This step is usually performed by using experimental data, simulations of previous models or relying on experts. Moreover, it is important to clearly identify the uses of the model being built. Determine what is to be predicted and under which assumptions. In some cases, experiments must be planned to generate the missing data. It is important to assure that the data covers the range of validity of the model. If experiments can be well controlled and do not take too long to be performed, it may be reasonable to consider generating a set of data for model validation as well, as in the Holdout validation approach, presented in 2.4.2.

After gathering the data, if it comes from experimental results, it is important to analyze it in order to remove eventual outliers and to check the level of variability. This helps setting the goals of the optimization, i.e., a model cannot be more precise then the noise level present in the data unless complementary information about the noise is taken into account.

Once the data is already evaluated and the target of the model is well defined, a set of candidate models may be proposed. For this purpose it is usual to plot the data into several charts, trying to identify relations among parameters. Once the first candidate models are ready, it is possible to obtain their parameters by optimizing the created equations to the experimental results, as presented in section 2.2. This step relies in a verified implementation of the model, as discussed in section 2.3.

The next step should be to validate the generated models. If specific data was generated for validation it may be used but, in either case, a Cross-validation approach is always feasible and prevents the model from overfitting the data available, see section 2.4.1.

All the models that passed validation may now be evaluated and the best one may be selected. Several selection criteria should be used together in order to ensure a good selection. One of the first approaches should take into consideration the cost function used for calibration, like an RMSE, presented in 2.5.1.1 and models presenting similar values could be distinguished by means of the AICc, presented in 2.5.2.1, plus the Adjusted R^2, presented in 2.5.2.3. Of course that if the data used for calibration and validation may still present outliers, the Interquartile Range, in section 2.5.1.4, is an interesting criterion to exclude some models from the selection list. Finally, depending on the use of the model, the Minimum and Maximum Errors, presented in section 2.5.1.3 (or their equivalent inside a Range, presented in 2.5.1.5) may be more relevant than the cost functions. This is clear the case of models where the extreme conditions are the most critical, such as in natural disaster or material resistance models.

If the model fails to pass the validation test, an improved version of it must be produced by either improving the description of the relation of the parameters that have impact on the phenomenon or by adding more parameters to the equations. The new generated models must then pass all the steps presented previously.

Chapter III. Sensitivity Analysis

"If I had eight hours to chop down a tree, I'd spend six hours sharpening my ax."

- *Abraham Lincoln*

3. Sensitivity Analysis

As shown in the previous section, one of the most important aspects of modeling is to identify which factors are relevant to the description of the phenomenon of interest. Nonetheless, most phenomena are influenced by a large number of factors. Besides, all factors are very unlikely to have the same importance. One of the strategies that may identify the most significant factors of a phenomenon is Sensitivity Analysis (SA) [SALTELLI, 2008], [IOOSS, 2015]. Moreover, SA may be used to understand the contributions and interactions among influence factors.

Generally speaking, "Sensitivity Analysis is the study of how the uncertainty, i.e. change of value, in the output of a model can be portioned to different sources of variation in the model inputs" [SALTELLI, 2004]. It mainly concerns the mathematical model representation of a physical system.

SA may be used for several purposes, such as decision making, quantification of a system or model development. For the purpose of this document, SA will be directed to Model Development tasks, as in model validation, simplification or calibration.

In this chapter the basic concepts of Sensitivity Analysis are presented and discussed. For the experimented reader it may be oversimplified or too restricted to the classical concepts. However, the approach of employing SA for lithography model evaluation, especially e-beam, is far from being widely employed and may help readers to follow the sequence of this work.

3.1. General Steps for Performing Sensitivity Analysis

Usually, any process of sensitivity analysis may be separated into five steps:

- Specify the right output variable for the problem;
- Select the inputs of interest;
- Explore the input space;
- Evaluate the model for each sample set of input values;
- Apply an estimator of sensitivity to the output variable.

3.1.1. Specify the right output variable for the problem

Although the output of a model may be any value (or function, or map), usual sensitivity analysis is performed considering a single value as output variable. This simplifies the task of evaluating the variability on the output value of any model. In the applications where the model output is not a single scalar but a vector or a function, sensitivity analysis may be performed by concentrating the output to one value, for example by evaluating the surface above a given threshold. In the case where the output is a function, sensitivity analysis could be performed based on its mean value, and so on. However, some authors have proposed specific approaches such as [CAMPBELL, 2006], [MARREL, 2011] and [GAMBOA, 2013] to produce first SA results on such problems without such oversimplification.

In the case where the output of a model is a map, sensitivity analysis may be performed by concentrating the output to one value, for example by evaluating the surface above a given threshold. In the case where the output is a function, sensitivity analysis could be performed based on its mean value, and so on.

3.1.2. Select the inputs of interest

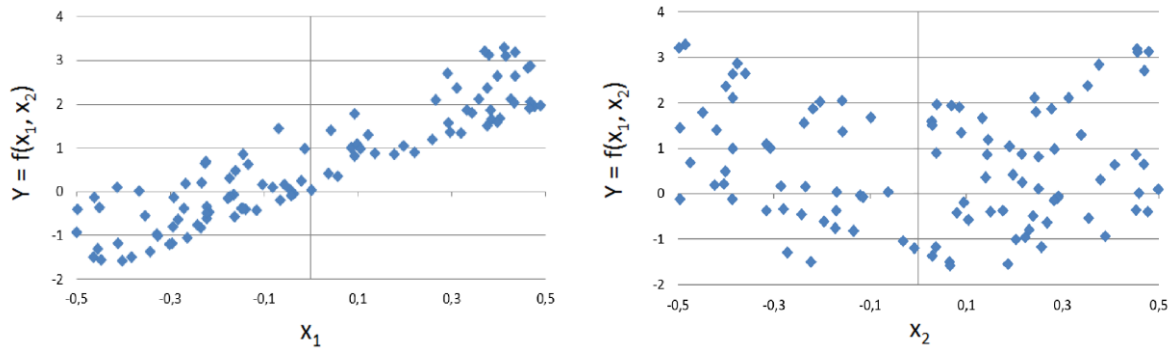
After determining which is the output variable, the second step is to determine what are actually the input variables and what are the elements that may be considered constant. Although each model input may be relevant, for models presenting tens or even hundreds of input variables it may be considered wise to narrow them down to a reasonable number. However, more important than reducing the number of inputs is to omit one input which could provide relevant information about the model behavior.

If the number of inputs under evaluation is overly reduced, the description of the problem is poor and may lead to an erroneous analysis. However, if the number of inputs under evaluation is overestimated, it might come to a point that no conclusion is possible. Ideally, every input selection should be justified in order to prevent the "Garbage In, Garbage Out" (GIGO) situation [SALTELLI, 2000]. GIGO is the expression used to reinforce the idea that a computer or any other logical process will unquestioningly process the data provided as input. If the input is unintended, nonsensical or false, the process will produce undesired and, often nonsensical output.

3.1.3. Explore the input space

The goal of this third step is to determine what range of the input set must be explored in order to evaluate the sensitivity of the output to the inputs. Moreover, the input uncertainty must be taken into account. If a measurement is considered as an input, for instance, the measurement error should be considered as the uncertainty of its value in the input (and this would be part of the interest of the sensitivity analysis). Characterizing the uncertainties is a challenge and may rely on the scientific literature, physical bounds, experiments, expert review and surveys, among others.

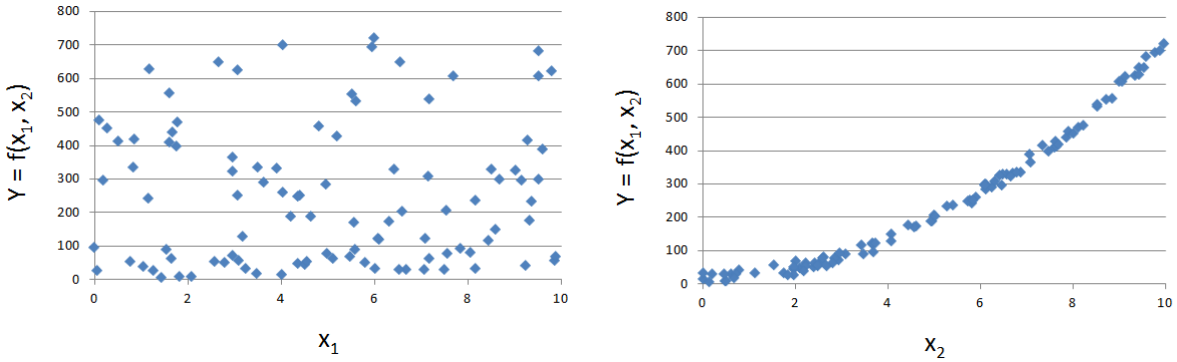
In order to better understand the importance of the input range to be determined, consider the following example: A given model is represented by a function $y = 4x_1 + 7x_2^2$. The question is which input parameter is more relevant, x_1 or x_2? A naive answer may consider that x_2 is more important since it is squared and has a larger coefficient. However, this answer cannot be provided until the ranges of variation of x_1 and x_2 are properly known. For instance, consider the parameters range [-0.5, 0.5] for both x_1 and x_2. The resulting graphs for one hundred random values of x_1 and x_2 generated inside the given range is presented in Figure 31.



(a) (b)

Figure 31. The result of 100 points randomly evaluation over the values of x_1 and x_2 in the range of [-0.5, 0.5]. (a) shows the model output as a function of x_1 while (b) shows the model output as a function of x_2.

Observe that Figure 31 (a) shows a clear influence of x_1 while Figure 31 (b) doesn't show such a clear trend for x_2. Based on this graph one may assume that x_1 has more impact than x_2. On the other hand, if the input ranges are changed from [-0.5, 0.5] to [0, 10], the results are completely different, as shown in Figure 32.



(a) (b)

Figure 32. The result of 100 points randomly evaluation over the values of x_1 and x_2 in the range of [0, 10]. (a) shows the model output as a function of x_1 while (b) shows the model output as a function of x_2.

Figure 32 shows a clear trend of the output of the model in respect of x_2 while x_1 seems to not have any impact over the result.

This is an important remark since it demonstrates that not only SA requires the relevant values to be inside the range under evaluation, but also that SA requires the range not to include values that are not likely at all to occur.

3.1.4. Evaluate the model for each sample set of input values

The next step is to evaluate the model for all input sets generated. One frequent assumption in this case is that the model is deterministic, i.e. for the same set of inputs it will always generate the same output. Non-deterministic models may be evaluated as well but, in this case, several runs for each input combination must be performed which makes the task more expensive. Once all computations of the model are finished, it is time to move to the final step.

3.1.5. Apply an estimator of sensitivity to the output variable

The last step of the process of sensitivity analysis is actually to evaluate the set of outputs corresponding to the sampled inputs. There are several algorithms for estimating the sensitivity and they are described in the next sections of this chapter.

Mainly, all sensitivity analysis techniques may be divided into two different categories: Local Sensitivity Analysis and Global Sensitivity Analysis [SALTELLI, 2008]. The differences between these two approaches impact mainly the steps presented in sections 3.1.3 and 3.1.5.

3.2. Local Sensitivity Analysis

Local Sensitivity Analysis is based on the idea of performing small perturbations in the inputs and observing its influence in the output of the model. Since the perturbations are small, only the influence in a small region is explored and that is the reason why this method is called local. One of the most popular methods of local sensitivity analysis is based in One-At-a-Time (OAT) factor changes. OAT technique consists in selecting a point in the input space and then changing the value of one of the coordinates of the input variable space. This procedure is repeated for each coordinate at a time, hence the name of the method, and conclusions are taken from this point. Since only one input value changes at a time, any change in the output can only be attributed to that particular input.

The accuracy of this approach strongly depends on the fact that the influence of one variable does not depend of the others, which is usually not the case.

For instance, consider the problem of estimating the surface of an unknown shape (which is actually a circle of radius 1) centered in zero, as shown in Figure 33. A circle was chosen to demonstrate the influence between variables using OAT techniques as its shape depends on both variables that are not independent. Consider that our reference point is the center point of the shape. Starting from it and varying only the X coordinate value, one will identify the shape limits in the Y coordinate are located at -1 and 1. Performing the same evaluation but this time only varying the Y coordinate, the limits of X would be reached in -1 and 1. Without the ability of varying both coordinates at the same time, one may take the shape for a square instead of a circle (as shown in Figure 33). Therefore, the conclusion of this method would be that the surface of the unknown shape under evaluation is the surface of a rectangle of width equal to 2 and height equal to 2. The error in precision, on this case is 3.14 / 4 = 0.785. This error gets more important as the number of dimensions of the space grows.

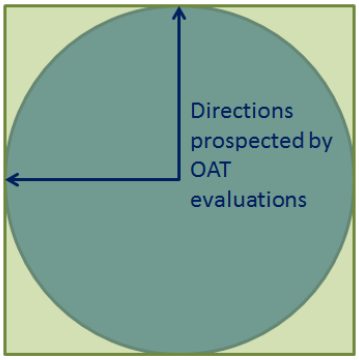


Figure 33. The OAT estimation over a circle, estimated as a square.

Another weakness of this approach is the influence of the initial point selected to perform the evaluation on the final shape using OAT. For instance, check four different rectangles representing the different area estimations of the circle from the previous example in the OAT approach according to four different initial point selection (Figure 34).

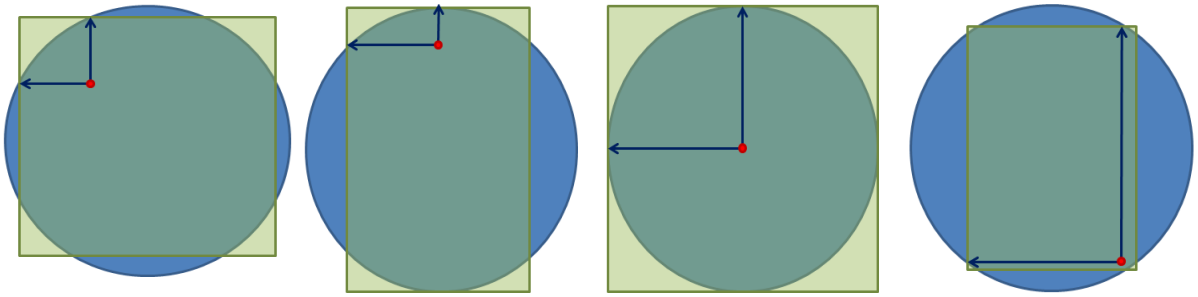


Figure 34. OAT estimation over a circle for four different starting point positions. For each starting point, a different estimation result is presented as a rectangle.

Therefore, OAT itself is not much useful in more complex spaces (dependent variables or high dimensions) and is usually considered a useful approach only for obtaining a first impression of the model behavior or when the model is so complex that more sophisticated approaches are not feasible.

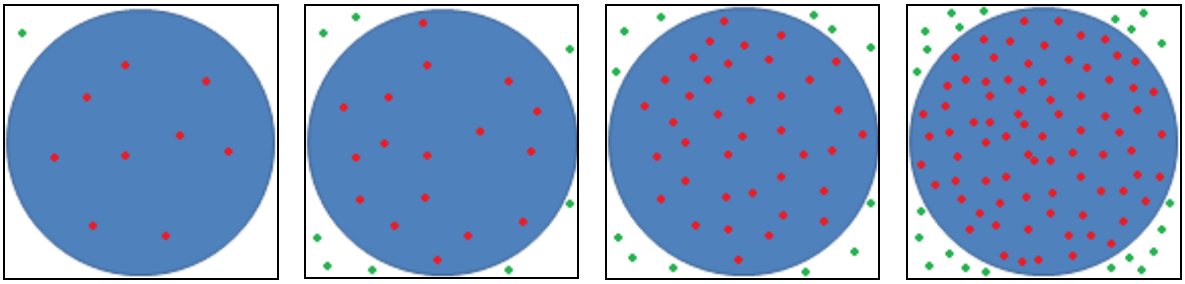
In order to better explore the input's space, another category of sensitivity analysis algorithms exists and is called Global Sensitivity Analysis.

3.3. Global Sensitivity Analysis

Global Sensitivity Analysis is based on the idea of sampling the input's space in a uniform and uncorrelated manner in order to evaluate the impact of several changes on the inputs at once. In comparison with the Local Sensitivity Analysis approach the main difference is the absence of a starting point. Instead, a set of points is evaluated.

Once again consider the problem of estimating the surface of a shape (which is the same circle of unitary radius previously presented in section 3.2). Consider the same input range variation and center position as the previous case. Considering the estimation of the shape surface based on a global analysis, one could imagine points randomly sampled inside the input ranges and

flagged as either inside or outside the shape. The ratio between the number of points inside the shape and the total number of points provide an estimation of the surface. The area of this circle is π, which in relative perspective represents 78.5% of the input space range (which is a square of size 2). Figure 35 shows the results using a sample size of (a) 10 points, (b) 25 points, (c) 50 points and (d) 100 points in a random sampling approach.



(a) (b) (c) (d)

Figure 35. Global Sensitivity Analysis estimation over a circle for four different number of point on the sample set.

For the first case, 9 points out of 10 are found inside the circle. On this case the total area would be estimated as 90% of the input ranges. For the second case, 17 out of 25 are found inside the circle, changing the estimation to 68%. Finally, when using 50 points the estimation goes to 76% while for 100 points the estimated are is 78% of the input ranges. As the number of points increases, the result should converge to the exact value of 78.5%.

For the implementation of a global sensitivity analysis approach, it is necessary to determine a statistical distribution to each input and then to sample the inputs. This is discussed in the next section.

3.3.1. Generate a sample of the inputs distributions:

After determining the range of each significant input of the model, it is important to select which points of this n-dimensional space must be sampled (and evaluated). Therefore, selecting a technique that obtains a significant coverage of the input space with a reduced number of points is of primary importance at this step. The result of a poor sampling of the space may be either the underestimation or the overestimation of the impact of a given parameter to the output value.

There are several different techniques for sampling the space of the inputs and this topic is covered in section 3.4.

3.4. Sampling Techniques

One of the major aspects that have an important influence over the results of any Global Sensitivity Analysis algorithm is the values used for the evaluation of the inputs. It is important to obtain a good sampling of the input space so a general view of the input combinations is taken into account in order to estimate their influence over the results of the model.

In order to evaluate and compare different sampling techniques, one must determine what defines a good sampling. There are different criteria and the choice of the one to be use depends on the application and how critical the sampling is for the purpose it is intended to. When it comes to sampling for Sensitivity Analysis, it is relevant to identify the sampling points that better cover the input space [MARELL, 2008]. In general, a good coverage of the input space is associated to covering the largest area possible in an as uniform as possible manner. However, non-uniformity may be desirable if different regions of the space present different importance.

Section 3.4.1 discusses some of the different criteria that may be used to evaluate the quality of a sampling.

3.4.1. Sampling Evaluation Criteria

The general goal of sampling for sensitivity analysis is to optimize the balance between number of points used and the input space coverage achieved by these points. Space coverage is the characteristic to cover the largest region in the space possible with the finest degree possible. The sampling result should provide information about all portions of the input parameter space if there is no prior knowledge about some regions. In some specific cases, this rule may be broken by a biased sampling but, in the context of Sensitivity Analysis, it is relevant to search the best space coverage possible [MARELL, 2008].

There are different metrics that provide information about the space coverage. They may rely in different properties as the density of points in a given region, the projection over each dimension or the distance between the points. Some of the most frequently used metrics are presented in the next sections.

3.4.1.1. Discrepancy

The discrepancy criterion evaluates the uniformity of the distribution in the space [FANG, 2001], [JIN 2005], [FEUILLARD, 2007]. The discrepancy is the deviation of the repartition of the points of a given sample from a uniform sample. Therefore, the smaller the discrepancy, the better the sampling covers the input space uniformly.

A formal definition of the discrepancy is presented in (30) [FAURE, 2012].

$$D^*(x) = \sup_{y \in [0,1)^k} \left| \frac{1}{N} \sum_{k=1}^{N} 1_{\{x^{(k)} \in [0,y]\}} - Volume\,([0,y]) \right| \quad (30)$$

A simple interpretation of this formula is the comparison of the volume of each sub-interval of the domain x ($Volume\,([0,y])$) and the proportion of points present in this sub-interval ($\frac{1}{N}\sum_{k=1}^{N} 1_{\{x^{(k)} \in [0,y]\}}$). The supremum norm allows to quantify the maximum difference for a given set.

Unfortunately this definition is not simple to compute due to the $L\infty$ from the supremum norm used on its definition. For convenience, a different norm calculation is used L^2, which gives the discrepancy described in (31).

$$D_2^*(x) = \left\{ \int_{[0,1)^k} \frac{1}{N} \sum_{k=1}^{N} 1_{\{x^{(k)} \in [0,y]\}} - Volume\ ([0,y])^2 dy \right\}^{1/2} \quad (31)$$

Discrepancy measures based on L^2-norm are the most popular in practice because closed-form easy-to-compute expressions are available. Two of them are particularly interesting due to their robustness to the reduction of dimensions: the centered L^2-discrepancy [DAMBLIN, 2013] and the wrap-around L^2-discrepancy [HICKERNELL, 1998].

Although this approach provides directly the information of the occupation of the input space, it presents the drawback of being expensive to compute, which makes it not the most attractive approach for evaluating spacing with a large number of dimensions.

3.4.1.2. Axis Projection

Another criterion for evaluating the quality of the space sampling coverage is to project the sampled points into each of the coordinate's axis and to determine how these projected points are distributed along the considered axis. This aspect is important for assuring a proper evaluation of the influence of each input. The blank spaces in an axis projection indicate input values that will not be evaluated at all. Overlapping points, i.e. points that are sampled in the same coordinate in one of the inputs, indicate that the input sampling coverage will be reduced and the sensitivity will be biased (or much more points will be required to achieve the same level of precision).

A sampling that presents its points uniformly distributed on the axis enhance the quality of the evaluations performed. The metric to axis projection uniformity may be the maximum distance between two samples, the minimum distance, the mean distance, the variance of the distances, etc.

Figure 36 presents three different sampling results and their projection to one of the input axis. Notice in Figure 36 (a) the same value is sampled several times for the given input coordinate while in Figure 36 (b) the lack of uniformity of the projection is the issue. Figure 36 (c) presents a more uniform projection on x_1.

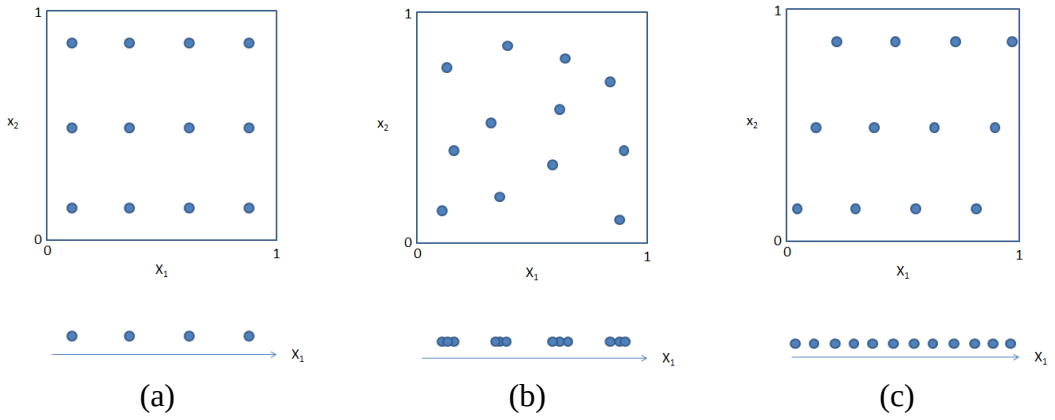
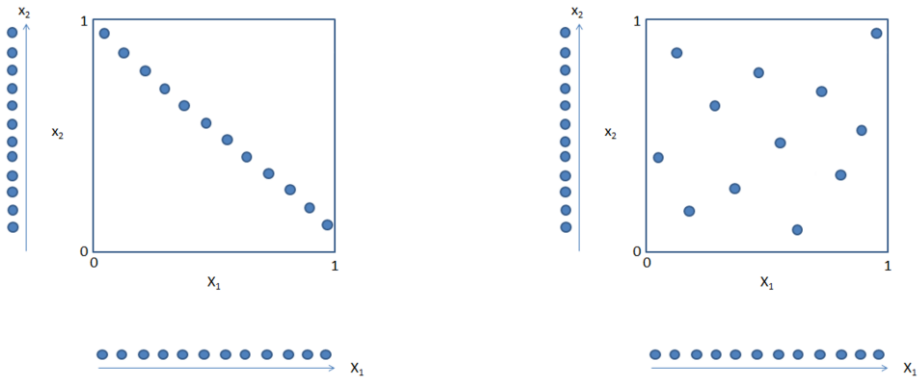


Figure 36. Three different sampling results and their projection to one of the axis.(a) shows an example of several points at the same coordinate; (b) shows an example of bad uniformity in the projection while (c) presents a more uniform projection.

The axis projection metric is a straightforward, fast and simple way to evaluate any sampling. However, although presenting a far from uniform axis projection in one or more inputs is a strong sign of a poor sampling, presenting a uniform one does not assure a good coverage of the space. One example of very uniform axis projection and yet of a poor coverage is shown in Figure 37 (a).



(a) (b)

Figure 37. Two examples of sampling presenting axis projections with high uniformity but one (a) with a poor space coverage and another (b) presenting a much better coverage.

Observe that better space coverage may produce projections identical to a worst space covered as shown in the comparison between Figure 37 (a) and (b).

3.4.1.3. Maximin

The Maximin index is a criterion to be maximized. It consists in taking the maximum of the minimum distances among all points of a sample [JOHNSON, 1990]. Ideally a sampling that presents a large minimum distance has adequate space coverage since there are no points too close to each other. For example, Figure 37 (a) and (b) samplings present the same projection on axis but the points in (a) are closer and thus (b) has a larger minimum distance. The weakness of this criterion is that it is not suitable to compare samples with different number of points since lower number of points tends to give higher (better) Maximin values.

3.4.1.4. Minimax

The Minimax index is a criterion to be minimized. It consists in taking the minimum of the maximum distances among all points of the sampling result [JOHNSON, 1990]. The reason is that if all points are overlapping each other on the center of the space, this criterion is minimal, and the space coverage is poor. Nonetheless, if the points are distributed in a given structure, this strategy may be useful. However, computing the Minimax value may be computationally expensive since all points must be visited and compared among themselves which, in the case a large dimensional space, may be prohibitive (the complexity is n^2). For this reason, Minimax is not a widely used criterion for sampling evaluation.

3.4.1.5. Maximal Volume

This criterion consists in taking the maximum of the volume occupied by hyperspheres centered in each point of the sampling result. It's a criterion that should be maximized.

The radius of each hypersphere is determined by half of the minimum distance between all points. It is mainly used in intermediate evaluations during the iterative construction of a sampling distribution [MARELL, 2008].

This metric is similar to the Maximin one in the sense that it relies on the minimal distance between two points. However, it is advantageous when comparing space sampling with different number of points. The Maximin criterion is usually negatively impacted by the increase of the number of points while the Maximal Volume may be either positively or negatively impacted and it is this tradeoff that may be used for determining the number of points on the sampling.

3.4.2. Sampling Strategies

There are several different strategies to distribute a certain number of points in order to sample an n-dimensional space. Some strategies are based on randomness using pseudo-random number generators while others are deterministic, assuring one or more interesting properties. In this chapter the most popular strategies are presented and discussed in terms of their strong and weak points.

3.4.2.1. Factorial Sampling

The first approach one can imagine to sample a space is to uniformly distribute samples. Each region of the space is divided into $\sqrt[k]{N}$ identical bins, where k is the number of dimensions of the space and N is the number of points. Centered on each k dimensional bin a sampling point is placed. This strategy is called Factorial Sampling. Figure 38 shows one example of this strategy for a 2 dimensional space input space sampled using 25 points.

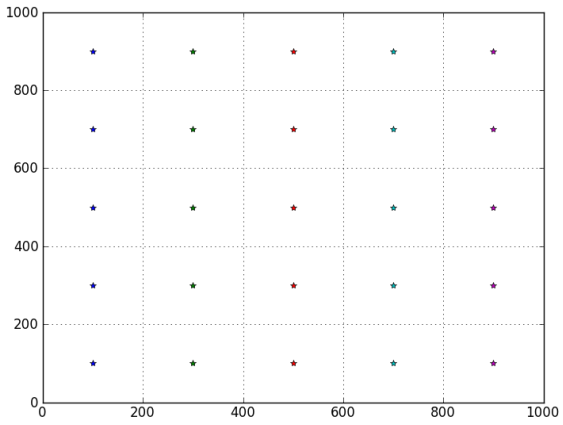


Figure 38. Example of Factorial Sampling with two inputs from 0 to 1000.

Notice that the space coverage is good in most of the metrics presented in the previous chapter. The only weakness is the projection on axis, due to the overlapping produced by the bins that are uniformly distributed. For instance, in a 6 dimension space, which would be necessary to evaluate a 6 input model, if 1,000,000 points are used, each axis projection will get only 10 points, which is far too low, especially considering the large number of points used.

3.4.2.2. Random Sampling

Another straightforward approach to sample is to randomly distribute the points over the input space. For that, a random value is determined for each input variable and the set of values determine the coordinate of a sample point. The process is repeated as many times as the number of points requires.

This approach is known as Monte Carlo Sampling. Figure 39 shows one example of this strategy for a 2D input space.

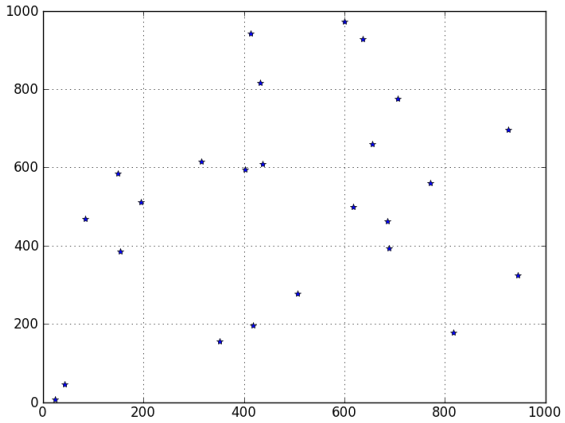


Figure 39. Example of Monte Carlo Sampling with two inputs from 0 to 1000.

Notice that this strategy provides a better projection over the axis than the Factorial Sampling approach. However, although it is unlikely that the projection of these samples to the axis will generate overlaps (since the random values are unlikely to be the same), some blanks (gaps) can still appear according to the number of points sampled.

Hence, counting on full randomness to provide a good coverage is far from ideal. The example presented in Figure 39 shows a visible lack of uniformity leaving some blanks and generating some clusters. This characteristic has a direct impact in the Discrepancy Index, which may vary significantly from one run to the next, but is rarely close to 0. Moreover, since this method is completely random, the sample quality may vary a lot, especially for low sampling number of points.

3.4.2.3. Latin Hypercube Sampling

In order to combine the strong points of the factorial approach (more uniform covering of the space) with the Monte Carlo approach (better axis projection), the Latin Hypercube Sampling (LHS) approach was proposed [MCKAY, 1979].

The goal of the Latin Hypercube Sampling is to ensure that for a given number of divisions over each input, every division gets the same number of samples. This structure may be achieved by the following:

- For each input, generate a vector containing as many elements as the number of points that will be used to sample the input space. For instance, consider a two inputs space

that should be sampled using 5 points. The two resulting vectors (one for each input variable) are presented below:

Vx_1: [0, 1, 2, 3, 4], Vx_2: [0, 1, 2, 3, 4]

- The second step is to randomly shuffle each vector. One of the possible resulting vectors is presented below:

$Vx_{1\ shuffled}$: [1, 3, 4, 0, 2], $Vx_{2\ shuffled}$: [3, 2, 1, 4, 0]

- The final step is to combine the vectors to generate the coordinates of each point. For the two vectors above, the five points would be the following (in the format P[x1, x2]):

P0 [1, 3], P1 [3, 2], P2 [4, 1], P3 [0, 4], P4 [2, 0]

The five points presented above are them distributed on the input space, which is divided in equally sized bins. For instance, P0 is set to the bin at position 1 in the X_1 coordinate and position 3 in the X_2 coordinate. The process is repeated for all points. The points may either be placed on the center of the correspondent bin or, in some variations, be randomly placed inside the bin. The result of this operation is shown in Figure 40.

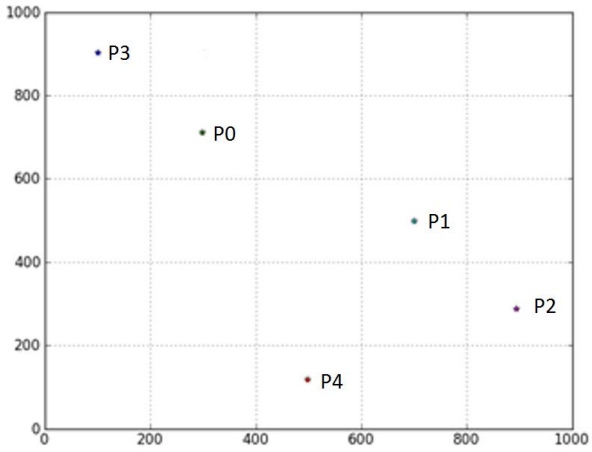


Figure 40. Example of LHS sampling giving in the step by step description of the algorithm.

Although this approach may combine the strengths of the factorial and the Monte Carlo approaches it, in certain way, also combine its weak points. If in one hand the factorial uniformity is lost, on the other hand the quality of the result may vary significantly from one execution to the next due to the randomness of the algorithm. One example using 25 points (and hence comparable with the ones presented in Figure 38 and Figure 39) is shown in Figure 41.

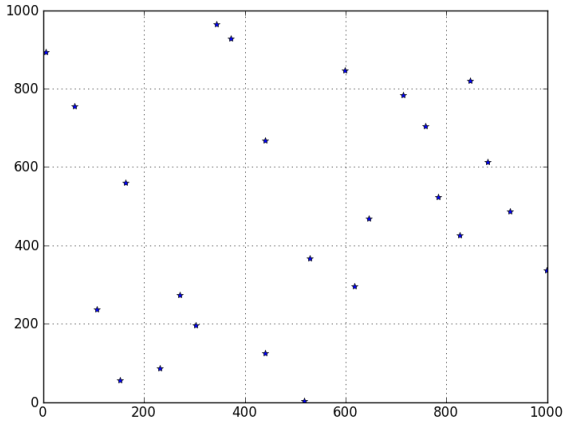


Figure 41. Example of LHS sampling with two inputs form 0 to 1000.

Notice that the space coverage is not ideal, but is better than Monte Carlo Sampling in most of the cases. The Discrepancy Index, which may vary from one run to the next, is rarely above the one presented by a Monte Carlo Sampling, but is never as good as from the Factorial Sampling.

In terms of the projection to the axis, LHS samples will generate no overlaps at all, since for each partition only one point is assigned. Actually, the spaces between any two points in the projections will be identical and will be of the size of each partition.

One drawback of this approach is that once the points are selected it is not possible to add more points while keeping the properties. The solution is to regenerate an entire sample with the new number of desired points. A simple random approach can continually be enriched by adding more points. However, it is possible to expect a better coverage using LHS even if less points are used. Finally, another issue is that it may present a bad coverage in some cases. For this purpose, some LHS variants have being proposed. Some of them are presented in the next sections.

3.4.2.4. U-Sampling

The U-Sampling Latin Hypercube sampling approach [TANG, 1993] proposes the use of Orthogonal Arrays as the initial version to present a better space coverage than the regular LHS.

Generally, an Orthogonal Array, or OA, is a matrix with *n* lines and *d* columns filled with *s* different integers. The restriction is that every sub-matrix of dimension *n* x *r* must present all line vectors 1 x *r* possible with exactly the same frequency. On this case, *r* is called the strength of the OA. For short, the syntax to represent one orthogonal array is OA(*n*, *d*, *s*, *r*). Consider one example of OA(9, 4, 3, 2) given in Figure 42 (a). Observe the three sub-matrices highlighted in Figure 42 (b) and the line vectors in each sub-matrix indicated by different colors (Figure 42(c)). Observe that each line vector occur one time in each sub-matrix, respecting the restriction described:

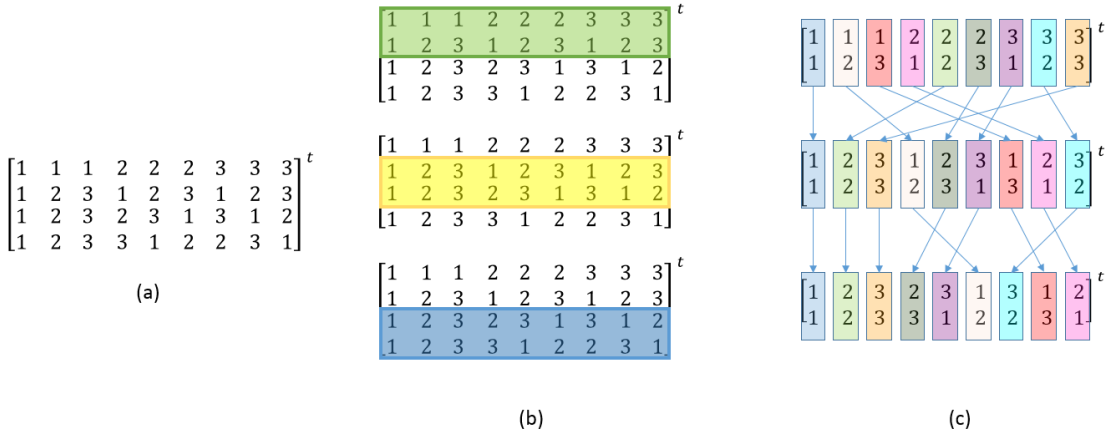


Figure 42. (a) Example of a OA(9,4,3,2); (b) the highlight of each of the sub-matrices and (c) highlight of each line vector of the sub-matrices.

A LHS may be represented by an OA(*n*, *d*, *s*, 1) where *n* is the number of sampling points distributed in *d* dimensions. However, as one may find several occurrences of the same number in the same column of an OA, it is necessary to replace the repeated numbers in any given order [TANG, 1991]. Using the example presented in [MARELL, 2008]:

$$\text{OA}(9, 2, 3, 2): \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \end{bmatrix}^t$$

For each column of the matrix (or lines in the transposed one), the repeated occurrences must be replaced. Since each number is repeated 3 times, the number 1 should be replaced by 1 in its first occurrence, by 2 in its second occurrence and by 3 in its third occurrence. The same replacement for numbers 2 (by 4, 5 and 6) and 3 (by 7, 8 and 9). The result of the replacements is presented below:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 4 & 7 & 2 & 5 & 8 & 3 & 6 & 9 \end{bmatrix}^t$$

If every line of the matrix above is taken as a set of coordinates x_1 and x_2 for building a LHS, the resulting sampling would be as presented in Figure 43.

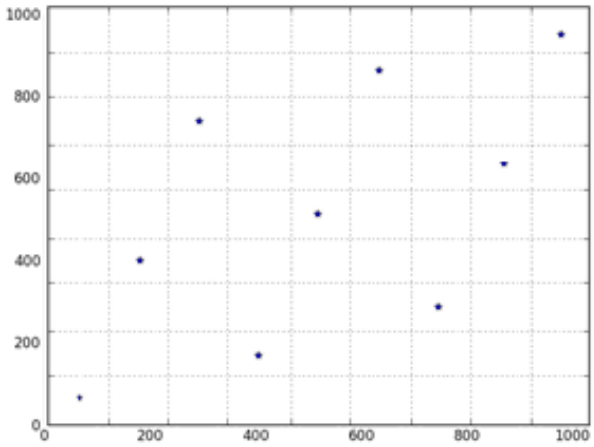


Figure 43. Example of U-sampling Latin Hypercube with two inputs form 0 to 1000.

Observe that the main advantage of this approach is that the space is divided into equally probable subspaces. This means that in any *r* dimension subspace will have the same number of sampling points, assuring a uniform coverage of the input space [TANG, 1993].

3.4.2.5. Orthogonal Latin Hypercube Sampling

The Orthogonal Latin Hypercube Sampling, OLHS, present its columns in an orthogonal form two by two by a scalar product [YE, 1998]. Although this algorithm does not present special advantages to space covering [YE, 1998], it generates a family of LHS which may be later selected according to an evaluation criteria, such as the ones presented in section 3.4.1. Moreover, another advantage of this approach is that the set of columns of the OLHS is completely uncorrelated [MARELL, 2008].

3.4.2.6. Iterative Generation of Random Samplings

For strategies presented in sections 3.4.2.2 and 3.4.2.3, as well as for any other random sampling strategy, it is possible to generate several realizations of the sampling and finally chose the best one, based on one or several criteria presented in section 3.4.1.

The evaluation criteria to be used should be fast to calculate in order to allow a large set of realizations to be computed. The Maximin criterion, presented in section 3.4.1.3, is one of the most frequently used strategies because it does not take a long time to evaluate, as oppose of Discrepancy or Minimax, which are less efficient. The strategy could be either to determine an initial number of realizations or to generate as many realizations as it takes until a given threshold value for the criterion index is reached. The advantage of performing several realizations is that the final quality is at least as good as the random strategy alone. The obvious disadvantage is that the computing cost is significantly higher than a single sampling.

3.4.2.7. Quasi-Random Sampling

Besides the factorial algorithm and all the random algorithms presented before, there are another class of algorithm known as Quasi-Random Sampling or Quasi-Monte

Carlo Sampling. These algorithms are deterministic and generate a highly uniform sampling based on strict rules for placing each point.

One popular Quasi-random approach to sample the space of inputs is known as Holton Low Discrepancy Sampling (LDS). This approach presents lower discrepancy than traditional LHS for small order of dimensions. However, for a large order, it has highly correlated projection.

Another popular Quasi-random sampling algorithm that also is classified as LDS is the Sobol' algorithm [SOBOL', 1967]. This approach may be seen as an improvement of the original Holton approach which, by respecting some given properties, produces uncorrelated samplings even in high dimensions. Those properties are almost equivalent to the strategies of LHS (based on dividing a space in blocks and ensure that at least one point is in each subcube).

Sobol' sampling presents the best uniformity of distribution as N goes to infinity, it is also a reasonable distribution for small initial sets and a very fast computational algorithm [MOROKOFF, 1994].

This algorithm is especially interesting when model evaluation takes a long time and the space must be well covered with as few points as possible. Moreover, it has the advantage of being constructive, which means that more points may be added at any time. However it is advisable to keep the number of points a power of two, $N = 2^m$, where m is an integer number, since the uniformity of the points distribution is only assured on these cases.

3.4.2.8. Sampling Strategies Comparison

In summary, there are many different sampling strategies and selecting one depends on the application and its constraints. In order to provide a view of some aspects that may be relevant – space coverage, projection to axis, computation cost and randomness – Table 7 is proposed. The randomness criterion accounts for how non-deterministic the algorithm is, meaning that running it several times will produce different results. The more ✓ signs a method presents, the better it is concerning that given criterion.

Table 7. Comparison among sampling strategies

Strategy	Coverage of the Space (Discrepancy)	Projection to Axis	Computation Cost	Randomness
Factorial	✓✓✓✓✓	✓	✓✓✓✓	✓
Monte Carlo	✓✓	✓✓	✓✓✓✓✓	✓✓✓✓✓
LHS	✓✓✓	✓✓✓	✓✓✓	✓✓✓
U-Sampling	✓✓✓✓	✓✓✓✓	✓✓	✓
Iterative LHS	✓✓✓✓	✓✓✓✓	✓	✓✓✓
Quasi-Random Sampling	✓✓✓	✓✓✓✓	✓✓	✓

It has been shown that, for a small number of points, a quasi-random sampling approach may provide the best coverage result in comparison to LHS based approaches, especially for a small number of dimensions. However, it has also being shown that, as the number of sampling points increase, even the Monte Carlo technique presents equivalent coverage. Therefore, employing a sophisticated approach such as the Quasi-Random Sampling that manage to obtain good coverage with a low discrepancy level and low correlation among any two inputs makes sense only if the sampling is limited to a small number of points – for instance, in the case where the evaluation of each point takes a long time. Moreover, even the most sophisticated Quasi-Random Sampling approach could not be demonstrated to be uncorrelated among all combinations of three inputs or more.

Another disadvantage of the LHS based approaches when compared to either a simple Monte Carlo approach or the Sobol' Quasi-Random approach is that it is not constructive - adding more points in a later moment do not maintain its main properties. Sobol' Quasi-Random algorithm may be used in a constructive fashion, but it is only uniform distributed when the total number of points sampled is a 2^n multiple. Monte Carlo runs are constructive since there is no relation between two points.

In summary, there is no particular approach that may be considered superior to all others in every aspect. For the application of SA for the purposes of this thesis, the selected approach was the Iterative LHS. The reason for this choice is that the model simulation effort is proportionally more expensive than the sampling cost, making this aspect less relevant than the others. Besides, this algorithm presents a good balance between the Coverage of the Space and the Projection to the Axis with the Randomness. This balance is less advantageous in the other approaches.

3.5. Global Sensitivity Analysis Algorithms

There are several different strategies to perform global sensitivity analysis. These strategies may be selected based on the analysis objective, the number of inputs of the model, its regularity and the computing time of a single simulation. Basically, there are three families of global sensitivity analysis methods, which are the regression methods, the screening methods and the variance-based methods.

The Regression methods [CACUCI, 2003][SALTELLI, 2000], allows ranking the sensitivity among parameters based on the relative magnitude of the regression coefficients that work as an indicative of the amount of influence the parameter has on the model. Methods of the Regression family, such as the Standardized Regression Coefficients (SRC), are usually the less computation expensive of the three families of methods. They are simple to code and require no specific design for the input selection. On the other hand, they are only suitable for linear or quasi-linear models [SALTELLI, 2005]. Rank based regression, such as the one presented at Standardized Ranked Regression Coefficients (SRRC), are an interesting alternative to overcome the SRC limitation. However, SRRC requires the model to be

monotonic and the conclusions of this method is not easily translated to the original model [TARANTOLA, 2006].

Screening methods in general are based on performing a local sensitivity analysis in several different points of the input space. By performing a good screening (hence its name) of the input space, this method obtains a global behavior of the model. Screening methods, such as the Morris method [MORRIS, 1991], are easier to compute and less time consuming than the other global methods. Moreover, less analysis points are required in comparison to the variance-based approach, for instance. Nonetheless, it provides only qualitative measurements of sensitivity and requires specific design of the inputs [CAMPOLONGO, 2007]. Finally, it is unclear the way to interpret the importance of each index of sensitivity.

Variance-based methods consider the variance of the output values as a function of the variance of the input values (more details are given in section 3.6). Variance-based methods are usually the most precise ones and are able to provide clear interpretation of the sensitivity indices and of the interactions of the inputs of any order. Nonetheless, they require special design of the input space, are usually computationally expensive and they assume that all information about the output is represented by its variance.

In order to summarize the pros and cons of each of the presented strategies for performing sensitivity analysis, two illustrations are presented in Figure 44. The first diagram (a) represents the limitation of each approach in terms of the model simulation time and the number of model inputs. On the second illustration (b) the methods are placed on relation to the importance of the assumptions about the model behavior and the effort to perform analysis over the results from the sensitivity result given by each method. These illustrations are after the ones presented by Saltelli [SALTELLI, 2005].

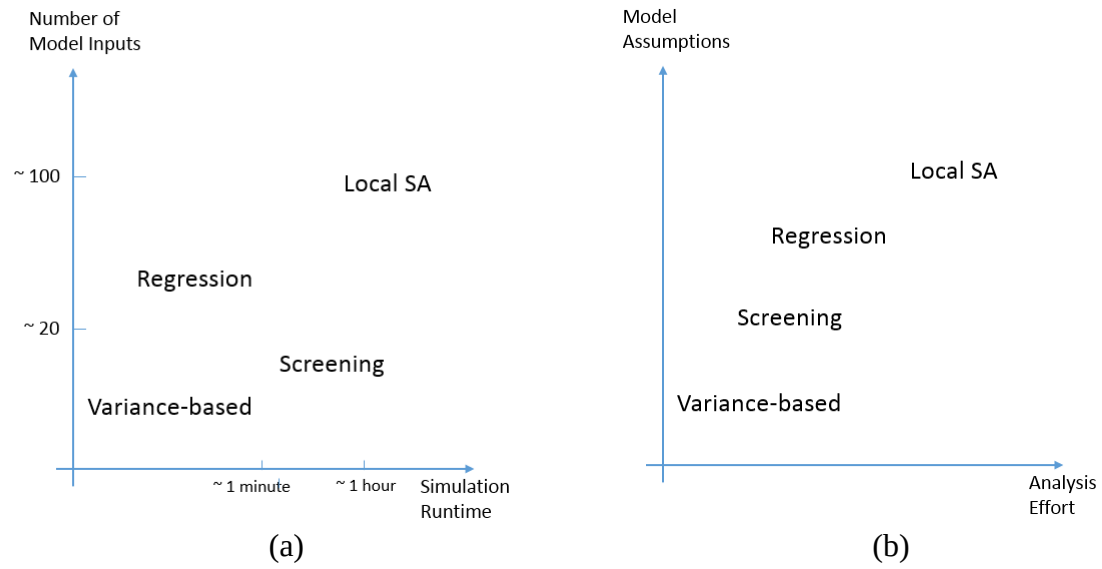


Figure 44. Illustration of factors to be considered while selecting a sensitivity analysis method (modified from [Saltelli 2005])

Observe that variance-based sensitivity analysis requires very little assumptions concerning the model and its results is straightforward but it is not suitable for models with large number of

inputs or long runtime per simulation. On the other hand local SA may be employed whenever the simulation runtime or the number of inputs are too high, but leading to higher analysis effort and imposing more model assumptions (such as monotonicity, linearity, etc.) in order to provide acceptable results.

In addition to these three families of global SA and to the local strategies described earlier, there is yet another approach that relies in metamodeling. This strategy consists in replacing an expensive model by simpler functions that provide similar results [VILLA-VIALANEIX, 2012]. Once this simplified model is identified, one may employ any of the previously mentioned sensitivity analysis methods over it. Notice that metamodeling requires selecting a function to replace the original model and to calibrate the parameters of such function.

In the context of compact models for lithography processes, they present in most cases a number of inputs smaller than 20 and a run time around a few seconds or less, depending on what is under simulation. Therefore, in the work presented in this thesis, the approach of choice is the Variance-based method, which is further explained in the next section.

3.6. Variance-based Sensitivity Analysis

Variance-based sensitivity analysis is one of the most popular forms of global sensitivity analysis. It is based on the decomposition of the variance of the output of a model into several contributor factors which are apportioned to each of the inputs of the model or to a combination of inputs.

As it was presented in the previews chapter, one may describe a generic model as Y = *f*(X), Y being the model output and X the input. Since variance-based sensitivity analysis handles the evaluation of a single output, if a model presents several outputs, sensitivity analysis may be performed independently for each output. In the following of this chapter, a model is consider to present a single output (i.e. Y = [y_1]) unless stated otherwise.

If it is assumed that all inputs are independent and that their distribution is uniform, it is possible to decompose the model function f(X) as presented in (32). This decomposition is called High Dimensional Model Representation (HDMR).

$$Y = f(X) = f_0 + \sum_{k=1}^{d} f_k(x_k) + \sum_{k<j}^{d} f_{k,j}(x_k, x_j) + \cdots + f_{1,2,\ldots,d}(x_1, x_2, \ldots, x_d) \quad (32)$$

Where f_0 is a constant, f_k is a function of x_k only, $f_{k,j}$ is a function of x_k and x_j, etc.

In a three dimensions example, the equation (32) could be written as in (33).

$$f(X) = f_0 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_{1,2}(x_1, x_2) + f_{1,3}(x_1, x_3) + f_{2,3}(x_2, x_3) + f_{1,2,3}(x_1, x_2, x_3) \quad (33)$$

There are many different ways to write a HDMR of a function. It has been proved that, if each term in the HDMR has zero mean, then all terms are orthogonal in pairs. As a consequence, these terms can be calculated using the conditional expectations of the model output.

$$f_0 = E(Y) \tag{34}$$

$$f_k(x_k) = E(Y|x_k) - f_0 \tag{35}$$

$$f_{k,j}(x_k, x_j) = E(Y|x_k, x_j) - f_0 - f_k - f_j \tag{36}$$

Where E(Y) is the mathematical expectation for Y, $E(Y|x_k)$ is the mathematical expectation given that Y as a function only of x_k and etc.

Based on these definitions, it is possible to observe that f_k is the effect of varying x_k alone, that $f_{k,j}$ is the effect of varying x_k and x_j discounted the independent factors f_k and f_j, and so on. By respecting these definitions, the HDMR is unique.

With the additional assumption that f(X) is square integrable, equation (32) may be squared and integrated to provide (37).

$$\int f^2(X)dX - {f_0}^2 = \sum_{k=1}^{d} \sum_{j_1<\cdots<j_k}^{d} \int {f^2}_{j_1 \ldots j_k}(x_k, x_j)\, dx_{j_1} \ldots dx_{j_k} \tag{37}$$

Equation (38) shows the result of squaring and integrating a three dimensions example, such as the one shown in equation (33).

$$\begin{aligned} \int f^2(X)dX - {f_0}^2 &= \int {f_1}^2(x_1)dx_1 + \int {f_2}^2(x_2)dx_2 + \int {f_3}^2(x_3)dx_3 \\ &+ \int\int {f_{1,2}}^2(x_1, x_2)dx_1dx_2 + \int\int {f_{1,3}}^2(x_1, x_3)dx_1dx_3 \\ &+ \int\int {f_{2,3}}^2(x_2, x_3)dx_2dx_3 + \int\int {f_{1,2,3}}^2(x_1, x_2, x_3)dx_1dx_2dx_3 \end{aligned} \tag{38}$$

Observe that, from equation (37) - or equation (38) in the three dimensions example - and the definition of variance, it is possible to write the generic equation (39) or its three dimensions example (40).

$$Var(f(X)) = \sum_{k=1}^{d} Var(f_k(x_k)) + \sum_{k<j}^{d} Var(f_{k,j}(x_k, x_j)) + \cdots + Var(f_{1,2,\ldots,d}(x_1, x_2, \ldots, x_d)) \tag{39}$$

$$\begin{aligned} Var(f(X)) = {} & Var(f_1(x_1)) + Var(f_2(x_2)) + Var(f_3(x_3)) \\ & + Var(f_{1,2}(x_1, x_2)) + Var(f_{1,3}(x_1, x_3)) + Var(f_{2,3}(x_2, x_3)) \\ & + Var(f_{1,2,3}(x_1, x_2, x_3)) \end{aligned} \tag{40}$$

These equations are expressed in a generic form (see equation (41) and (42)) for convenience.

$$V = \sum_{k=1}^{d} V_k + \sum_{k<j}^{d} V_{k,j} + \cdots + V_{1,2,\ldots,d} \tag{41}$$

$$V = V_1 + V_2 + V_3 + V_{1,2} + V_{1,3} + V_{2,3} + V_{1,2,3} \tag{42}$$

In this way, the variance of the output is the sum of the variance of all its components, and for all orders. The next step is to evaluate the relative importance of each input. This is usually done by computing indices, as it will be described in the following sections.

3.6.1. First-order indices

The decomposition presented in the previous section presents variance terms that are related to a unique input variable V_k. The portion of the total variance V apportioned to V_k alone is of first order. Based on these first order elements, it is possible to calculate their proportion compared to the total variance, as shown in (43).

$$S_k = \frac{V_k}{V} \tag{43}$$

Where S_k is called the first-order sensitivity index. This sensitivity index measures the effect of varying x_k alone, normalized by the total variance. Notice that from (41) and (43) the summation of all first-order indices is less than or equal to one (44).

$$\sum_{k=1}^{d} S_k \leq 1 \tag{44}$$

If the sum of all first-order indices equals one it means that there is no interaction between the inputs and the model may be considered additive. On the other hand, the closer to zero the summation is, the more important are the higher-order factors.

3.6.2. Higher-order indices

As stated in the previous section, the relevance of the interactions of the different inputs of a model may vary. It depends considerably of the model structure itself and the range of values considered for each of the inputs. Higher-order indices may be calculated in a similar way of the first-order indices, as shown for second-order in equation (45).

$$S_{kj} = \frac{V_{k,j}}{V} \tag{45}$$

3.6.3. Total-order indices

As the number of inputs grows, more indices may be required to understand the relationship among the inputs of a model. If all inputs interact, a *d* inputs model will require 2^d-1 indices of all orders. In this sense it was created the total-order index for a given input, S_{Tk}, which takes into consideration all the sensitivity indices concerning this input and which is calculated as shown in (46).

$$S_{T_k} = S_k + \sum_{j=1}^{d-1} S_{k,j} + \sum_{j=1}^{d-1} \sum_{l<l}^{d-2} S_{k,j,l} + \cdots + S_{1,2,\ldots,d} \tag{46}$$

Another way to compute the total-order index is based on the equation (47).

$$S_{T_k} = \frac{\mathrm{E}(\mathrm{V}(\mathrm{Y}|\mathrm{x}_{\sim \mathrm{k}}))}{V} \tag{47}$$

Where $\mathrm{E}(\mathrm{V}(\mathrm{Y}|\mathrm{x}_{\sim \mathrm{k}}))$is the mathematical expectation given that Y is a function of all variables but x_k.

Observe that while the summation of first-order indices should always be less than or equal to one, the summation of total-order indices are always bigger than or equal to one (48). If the summation of all total-order sensitivity indices is equal to one there are no interactions between the inputs and the model is completely additive.

$$\sum_{k=1}^{d} S_{T_k} \geq 1 \tag{48}$$

The first-order indices and the total-order indices are the sensitivity indices usually considered in variance-based sensitivity analysis. The total-order indices provide information about the impact of a variable in the model output variance while the first-order indices provide their impact without interactions with other inputs.

3.6.4. Example of Sensitivity Indices

In order to illustrate the concept of variance based sensitivity analysis presented so far let us consider the example presented at the beginning of the chapter. The model is described by the equation $y = 4x_1 + 7x_2^2$, with the values for inputs x_1 and x_2 on the range [-0.5, 0.5].

$f_0 = E(Y) = \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} (4x_1 + 7x_2^{\,2})\, dx_1 dx_2 = \frac{7}{12} = 0.5833$

$f_1(x_1) = E(Y|x_1) - f_0 = \int_{-0.5}^{0.5} (4x_1 + 7x_2^{\,2})\, dx_2 - \frac{7}{12} = 4x_1 + \frac{7}{12} - \frac{7}{12} = 4x_1$

$f_2(x_2) = E(Y|x_2) - f_0 = \int_{-0.5}^{0.5} (4x_1 + 7x_2^{\,2})\, dx_1 - \frac{7}{12} = 7x_2^{\,2} - \frac{7}{12}$

$f_{1,2}(x_1, x_2) = E(Y|x_1, x_2) - f_0 - f_k - f_j = 4x_1 + 7x_2^{\,2} - \frac{7}{12} - 4x_1 - \left(7x_2^{\,2} - \frac{7}{12}\right) = 0$

Therefore this HDMR respects the definitions presented earlier and is a unique solution. Observe that $f_{1,2}$ equals zero, which means that there is no interaction among x_1 and x_2 and only first order indices need to be computed. Then, it is possible to calculate each term of the variance of the function.

$$V_1 = \int f_1^2(x_1)dx_1 = \int_{-0.5}^{0.5}(4x_1)^2 dx_1 = 1.3333$$

$$V_2 = \int f_2^2(x_2)dx_2 = \int_{-0.5}^{0.5}(7x_2^2 - \frac{7}{12})^2 dx_2 = 0.2722$$

$$V_{12} = 0$$

$$V = V_1 + V_2 + V_{12} = 1.3333 + 0.2722 + 0 = 1.6056$$

From the definition shown in section 3.6.1, it is possible to determine the sensitivity indices for each input variable:

$$S_1 = \frac{V_1}{V} = \frac{1.3333}{1.6056} = 0.8304$$

$$S_2 = \frac{V_2}{V} = \frac{0.2722}{1.6056} = 0.1696$$

The sum of S_1 and S_2 equals 1, demonstrating that the model is fully additive.

If you look again to the plots shown in Figure 31, the calculation presented here confirms the impression that x_1 has more impact than x_2. Moreover, it quantifies the impact of each input by stating that the input x_1 accounts for around 83% of the variance in the output while x_2 only accounts for around 17%.

Unfortunately, such calculation is not always feasible. Sometimes it is not possible to directly describe the model as an integrable function, which is required to compute the variances and, hence, for the sensitivity indices calculation. Whenever an exact calculation is not feasible, one may rely in different strategies to provide estimations of the sensitivity indices. This topic is the subject of the next section.

3.6.5. Estimation of the Sensitivity Indices

The strategies are based on varying the values of all inputs and identifying the impact over the variance of the output. The first approach is based on scatterplots and is presented in 3.6.5.1 while a more complex and advantageous approach is presented in 3.6.5.2.

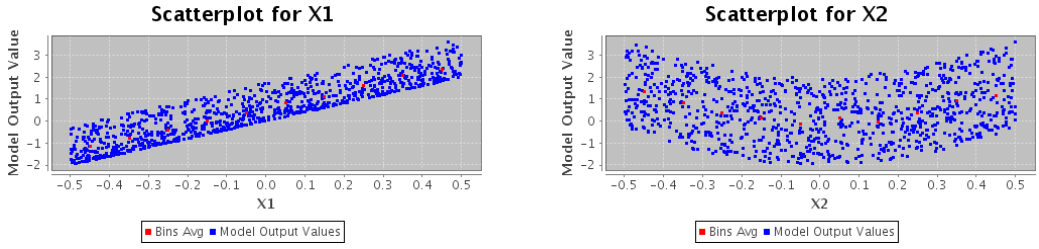
3.6.5.1. Sensitivity Indices Estimation based on Scatterplots

This approach consists in evaluating the output variance as a function of one input variable at a time. The steps to perform in order to run the estimation based in scatterplots are the following:

- Several instances of the model are created, varying the model input parameter values in order to cover the model input space;

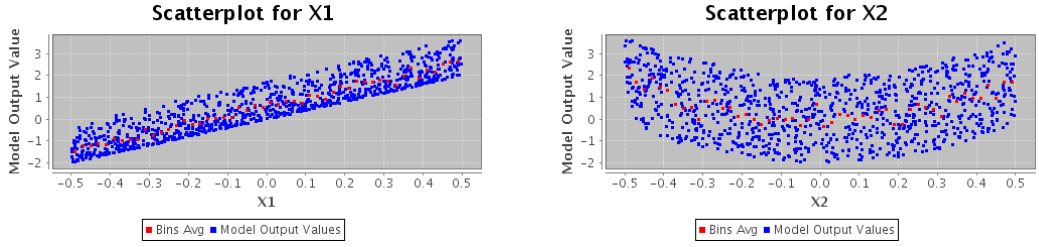
- The output value of the model is computed for each of the previously generated input set;
- A scatterplot is built for each input variable. It consists in plotting the output values as a function of the input under evaluation;
- Each scatterplot is divided into a certain number of bins, and the mean of output values inside each bin is calculated;
- The final step is to compute the sensitivity index for each input variable as the ration between the variance among the bins and the total variance of the sampled points.

As an example, consider the plots presented in Figure 31. Scatterplots are divided into a given number of bins and the mean of each bin is calculated. Here, the mean value is represented as a red point centered on the bin region. Observe in Figure 45 the result of dividing a sample of 1000 points into 10 bins, in Figure 46 the result for 1000 into 50 bins and in Figure 47 and Figure 48 the division of 64000 points into 10 and 50 bins, respectively.



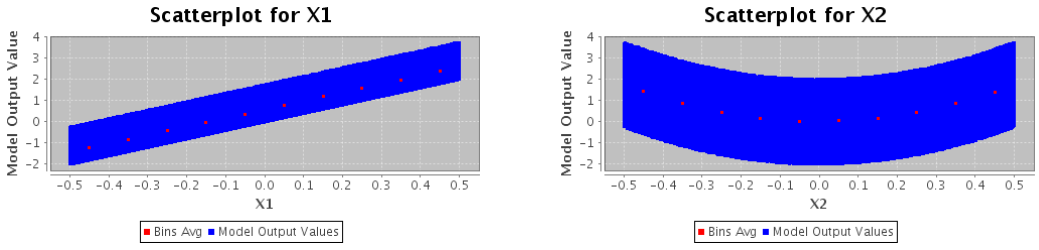
(a) (b)

Figure 45. The result of 1000 points randomly evaluation over the values of (a) x1 and (b) x2 in the range of [-0.5, 0.5] and the 10 bin averages.



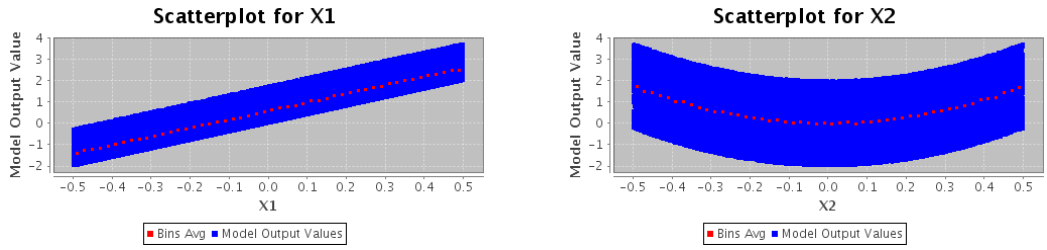
(a) (b)

Figure 46. The result of 1000 points randomly evaluation over the values of (a) x1 and (b) x2 in the range of [-0.5, 0.5] and the 50 bin averages.



(a) (b)

Figure 47. The result of 64000 points randomly evaluation over the values of (a) x1 and (b) x2 in the range of [-0.5, 0.5] and the 10 bin averages.



(a) (b)

Figure 48. The result of 64000 points randomly evaluation over the values of (a) x1 and (b) x2 in the range of [-0.5, 0.5] and the 50 bin averages.

Once the average value inside each bin is computed, the sensitivity index is calculated by dividing the variance among the average values of each bin and the variance of all point in the sample, as shown in (49).

$$S_i = \frac{V(BinsAvg)_i}{V} \tag{49}$$

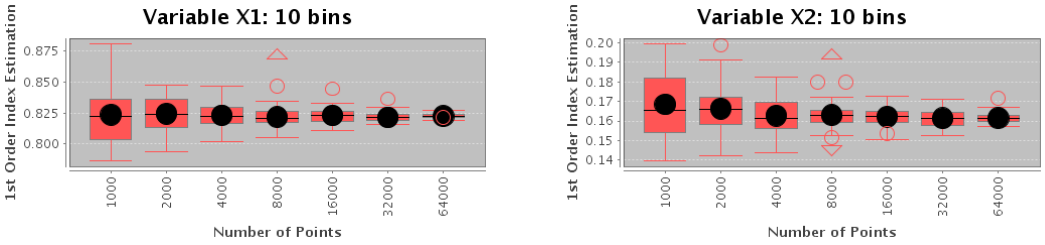
Estimated values of total-order sensitivity indices are given in Table 8 for the four previous cases. The reference values for the error calculation are S_{T1} and S_{T2} (0.83 and 0.17, respectively) previously calculated in 3.6.4.

Table 8. Example of estimation of first order indices based on scatterplots and the error associated for each case.

Example #	Figure	Points	Bins	S_1	S_{1err}	S_2	S_{2err}
1	Figure 45	1000	10	0.82	-0.01	0.16	-0.01
2	Figure 46	1000	50	0.85	0.02	0.25	0.08
3	Figure 47	64000	10	0.83	0	0.16	-0.01
4	Figure 48	64000	50	0.83	0	0.17	0

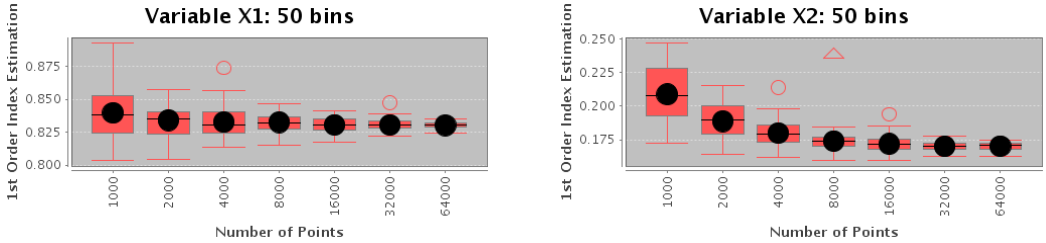
Looking on error on sensitivity index estimation in Table 8, one can see that the accuracy of this strategy is dependent of the size of the sampling and of the number of bins used. Observe the red dots (bin average) in Figure 46, especially in (b). From one point to the next there is a large variation. This impacts strongly on the error of S2 (0.08 in Table 8).

In order to evaluate the impact of increasing the number of bins and the number of points in the precision, another test is proposed. The algorithm is performed for 1000, 2000, 4000, 8000, 16000, 32000 and 64000 points and for 10, 50 and 100 bins, in all combinations. Each pair number of points/number of bins is performed 50 times, in order to evaluate the variability related to the randomness of the Monte Carlo sampling used. The results for X1 and X2 are presented in Figure 49, Figure 50 and Figure 51.



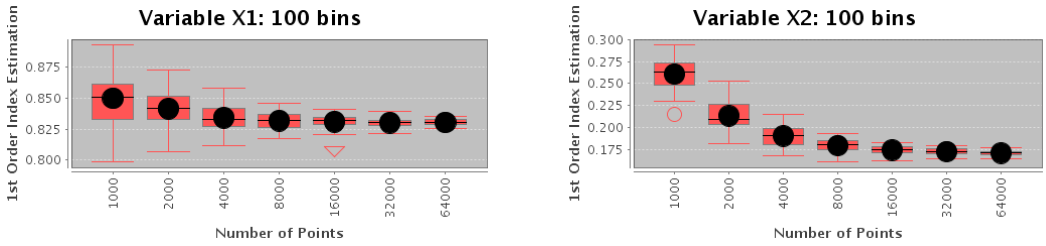
(a) (b)

Figure 49. The boxplot for (a) x1 and (b) x2 1st order estimation varying the number of points but using always 10 bins. Each box is calculated based on of 50 runs.



(a) (b)

Figure 50. The boxplot for (a) x_1 and (b) x_2 1st order estimation varying the number of points but using always 50 bins. Each box is calculated based on of 50 runs.



(a) (b)

Figure 51. The boxplot for (a) x1 and (b) x2 1st order estimation varying the number of points but using always 100 bins. Each box is calculated based on of 50 runs.

The boxplots presented above allow confirming the initial analysis of the scatterplot strategy. The first conclusion, which is not a surprise, is that the more points are used, the more precise is the result. The second conclusion is that each bin must contain a minimum number of points to be statistically representative for providing an accurate estimation of the local variance. It is observed that for a reduced number of points (such as 1000 points, for instance), the estimation tends to be closer to the analytical value while using 10 bins (around 100 points per bin) than for 100 bins (around 10 points per bin).

Nonetheless, this strategy presents a few drawbacks. First, it is suitable for determining only the total-order indices and, for this reason, may be less representative for models that present interaction among the inputs. The second drawback is the lack of precision if few points are used. For these reasons, another approach based on a design of experiment was proposed and is discussed in the next section.

3.6.5.2. Sensitivity Indices Estimation based on Saltelli/Sobol' method

Another strategy to estimate the sensitivity indices is based on estimating the variance values of a sample as local (OAT strategy), but performed several times. First, consider that the first order sensitivity index is calculated as shown in equation (43). V_k may be computed as follows:

$$\begin{aligned} V_k &= \int (\mathrm{E}(\mathrm{Y}|\mathrm{x}_k) - \mathrm{f}_0)^2 dx_1 \\ &= \int (\mathrm{E}(\mathrm{Y}|\mathrm{x}_k))^2 dx_1 - 2\mathrm{f}_0 \int \mathrm{E}(\mathrm{Y}|\mathrm{x}_k)\, dx_1 + \mathrm{f}_0{}^2 \\ &= \int (\mathrm{E}(\mathrm{Y}|\mathrm{x}_k))^2 dx_1 - \mathrm{f}_0{}^2 \end{aligned} \tag{50}$$

For estimating the $(\mathrm{E}(\mathrm{Y}|\mathrm{x}_k))^2$ term, consider two different sets of input variable values, given by X = [x_1, x_2, …, x_n] and X' = [x_1', x_2', …, x_n']. In order to estimate the first-order sensitivity index S_k for a given variable k, only the value of the given variable is kept constant, while all other variable values x_j (j≠k) vary. Therefore, for an example of three variables, considering the one under evaluation to be x_1 and the others x_2 and x_3:

$$\begin{aligned} (\mathrm{E}(\mathrm{Y}|\mathrm{x}_1))^2 &= \big(\mathrm{E}(\mathrm{Y}|\mathrm{x}_1)\big) * \big(\mathrm{E}(\mathrm{Y}|\mathrm{x}_1)\big) \\ &= \int f(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3)\, d\mathrm{x}_2 d\mathrm{x}_3 \\ &* \int f(\mathrm{x}_1, \mathrm{x}_2', \mathrm{x}_3')\, d\mathrm{x}_2' d\mathrm{x}_3' \end{aligned} \tag{51}$$

and its integral in respect of x_1

$$\int (\mathrm{E}(\mathrm{Y}|\mathrm{x}_k))^2 dx_1 = \int f(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3) f(\mathrm{x}_1, \mathrm{x}_2', \mathrm{x}_3')\, d\mathrm{x}_2 d\mathrm{x}_3 d\mathrm{x}_2' d\mathrm{x}_3' dx_1 \tag{52}$$

For the calculation of $\mathrm{f}_0{}^2$:

$$\mathrm{f}_0{}^2 = \big(\mathrm{E}(\mathrm{Y})\big)^2 = \big(\mathrm{E}(\mathrm{Y})\big) * \big(\mathrm{E}(\mathrm{Y})\big) = \int f(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3)\, d\mathrm{x}_1 d\mathrm{x}_2 d\mathrm{x}_3$$

$$* \int f(\mathrm{x}_1', \mathrm{x}_2', \mathrm{x}_3') d\mathrm{x}_1' d\mathrm{x}_2' d\mathrm{x}_3' \tag{53}$$

$$= \int f(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3) f(\mathrm{x}_1', \mathrm{x}_2', \mathrm{x}_3')\, d\mathrm{x}_1 d\mathrm{x}_2 d\mathrm{x}_3 d\mathrm{x}_1' d\mathrm{x}_2' d\mathrm{x}_3'$$

And by replacing (52) and (53) in (50) we obtain

$$\begin{aligned} V_1 = \int f(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3) * (f(\mathrm{x}_1, \mathrm{x}_2', \mathrm{x}_3') \\ - f(\mathrm{x}_1', \mathrm{x}_2', \mathrm{x}_3'))\, d\mathrm{x}_1 d\mathrm{x}_2 d\mathrm{x}_3 d\mathrm{x}_1' d\mathrm{x}_2' d\mathrm{x}_3' \end{aligned} \tag{54}$$

Equation (54) may be solved approximately by sampling several points on the input space using any of the techniques shown in section 3.4. For instance, for *N* points in the sample:

$$V_1 = \frac{1}{N-1}\sum_{r=1}^{n} f(x_{r1}, x_{r2}, x_{r3}) * (f(x_{r1}, x'_{r2}, x'_{r3}) - f(x_{r1}', x_{r2}', x_{r3}')) \quad (55)$$

Notice that for the three-dimensional example presented in this section, the model must be computed in five points so all three first-order parameters are computed. The points are $f(x_{r1}, x_{r2}, x_{r3})$ and $f(x'_{r1}, x'_{r2}, x'_{r3})$, which are used as reference in all calculations, plus the three points $f(x_{r1}, x'_{r2}, x'_{r3})$ for x_1, $f(x'_{r1}, x_{r2}, x'_{r3})$ for x_2 and $f(x'_{r1}, x'_{r2}, x_{r3})$ for x_3. For n-dimensional models, the number of points required is (n+2).

Finally, this strategy may also be used to compute the total order indices. For this purpose, initially consider the equation (47). The factor $E(V(Y|x_{\sim k}))$ may be estimated in a similar way as (55), as shown in (56).

$$E(V(Y|x_{\sim 1})) = \frac{1}{2N-1}\sum_{r=1}^{N} (f(x_{r1}, x_{r2}, x_{r3}) - (f(x'_{r1}, x_{r2}, x_{r3}))^2 \quad (56)$$

Notice that for the three-dimensional example presented in this section, the model must be computed in four points so all three total-order parameters are computed. The points are $f(x_{r1}, x_{r2}, x_{r3})$, which is used as reference in all calculations, plus the three opposite points $f(x'_{r1}, x_{r2}, x_{r3})$, for x_1, $f(x_{r1}, x'_{r2}, x_{r3})$, for x_2 and $f(x_{r1}, x_{r2}, x'_{r3})$, for x_3. For n-dimensional models, the number of points required is (n+1) [MARA, 2008].

Based on these two evaluations, the total number of points which a model should be computed to estimate first-order and total-order indices are (n+2) + (n+1) = (2n + 3), and, as there is one common point, 2*(n+1). Table 9 shows the points to be computed in order to estimate both the first-order and the total-order indices. The repeated points are represented in red.

Table 9. The points where the function must be evaluated for estimating the first-order and total-order indices. In red, the repeated points.

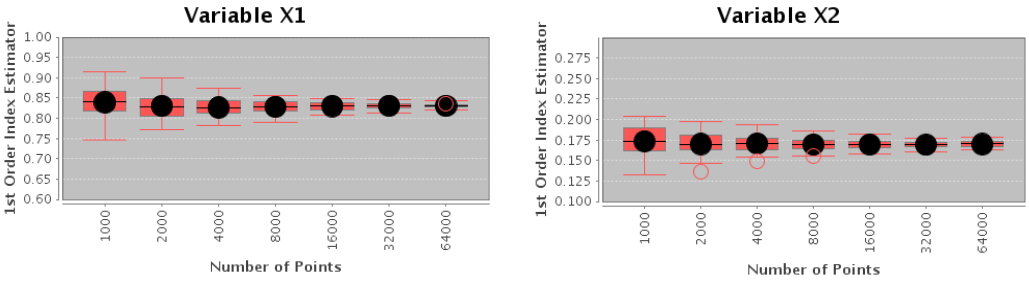
First-order indices	Total-order indices
$f(x_{r1}, x_{r2}, x_{r3})$	$f(x_{r1}, x_{r2}, x_{r3})$
$f(x_{r1}, x'_{r2}, x'_{r3})$	$f(x'_{r1}, x_{r2}, x_{r3})$
$f(x'_{r1}, x_{r2}, x'_{r3})$	$f(x_{r1}, x'_{r2}, x_{r3})$
$f(x'_{r1}, x'_{r2}, x_{r3})$	$f(x_{r1}, x_{r2}, x'_{r3})$
$f(x'_{r1}, x'_{r2}, x'_{r3})$	

However, by changing the reference point of the first-order calculation from X to X', this number may be reduced to (n+2) again for all indices together. Table 10 shows the points to be computed in order to estimate both the first-order and the total-order indices based on this changing of reference. The repeated points are represented in red.

Table 10. The points where the function must be evaluated for estimating the first-order and total-order indices based on the variation proposed by [MARA, 2008]. In red, the repeated points.

First-order indices	Total-order indices
$f(\mathbf{x'}_{r1}, \mathbf{x'}_{r2}, \mathbf{x'}_{r3})$	$f(\mathbf{x}_{r1}, \mathbf{x}_{r2}, \mathbf{x}_{r3})$
$f(\mathbf{x'}_{r1}, \mathbf{x}_{r2}, \mathbf{x}_{r3})$	$f(\mathbf{x'}_{r1}, \mathbf{x}_{r2}, \mathbf{x}_{r3})$
$f(\mathbf{x}_{r1}, \mathbf{x'}_{r2}, \mathbf{x}_{r3})$	$f(\mathbf{x}_{r1}, \mathbf{x'}_{r2}, \mathbf{x}_{r3})$
$f(\mathbf{x}_{r1}, \mathbf{x}_{r2}, \mathbf{x'}_{r3})$	$f(\mathbf{x}_{r1}, \mathbf{x}_{r2}, \mathbf{x'}_{r3})$
$f(\mathbf{x}_{r1}, \mathbf{x}_{r2}, \mathbf{x}_{r3})$	

In order to compare this approach with the scatterplot one, it was performed a test consisting in varying the number of points from 1000 to 64000 and running the algorithm 50 times, in a similar way to the test presented in Figure 49, Figure 50 and Figure 51. The result is shown in Figure 52.



(a) (b)

Figure 52. The boxplot for (a) x1 and (b) x2 1st order estimation varying the number of points. Each box is calculated based on 50 runs.

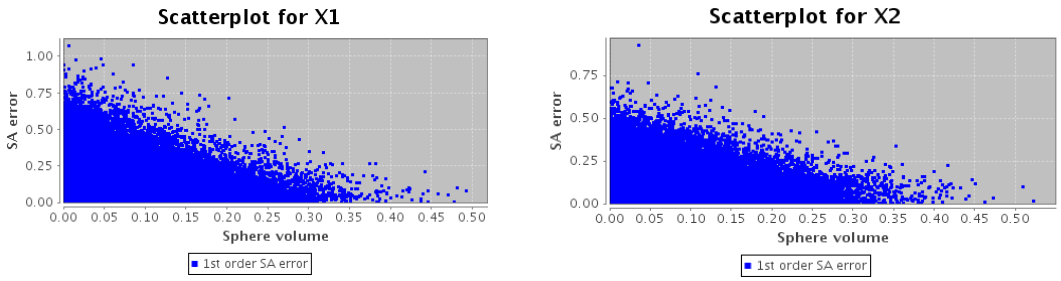
Compare the boxplot results presented in Figure 52 with the results from the scatterplots approach (presented in Figure 49, Figure 50 and Figure 51). The precision of this approach for the tests with small number of points (such as 1000 or 2000 points) is much better, and the deviation is smaller. Moreover, this approach allows computing both first-order and total-order indices, which is not the case for the scatterplots approach. For this reason, this is the strategy adopted in this work.

3.6.5.3. Importance of Sample Uniformity in Indices Estimation

In the sampling section it was stated that the more uniform the input space was sampled, the better the sensitivity analysis algorithm would perform. In order to verify this assumption, an example is proposed.

The test consists in generating 500000 different random samplings presenting only 50 points over the input space. Each sampling result is then evaluated in terms of space coverage. The technique used to evaluate each sampling result was the Maximal Volume (described in section 3.4.1.5).

For each sampling result, the Saltelli/Sobol' algorithm for estimating the sensitivity indices was performed. The analysis is to verify if there is a strong correlation between the error of the estimation and the quality of the sampling. The scatterplots showing the estimation errors for S_{T1} and S_{T2} are presented in value presented is the absolute difference between the theoretical value and the estimated one. The plots are presented in Figure 53.



(a) (b)

Figure 53. The result of 500000 runs for 50 points randomly estimation based on scatterplots over the values of (a) x1 and (b) x2 in the range of [-0.5, 0.5], using only 10 bins.

Observe that there is a clear correlation between the quality of the sampling (indicated by the volume occupied by the spheres) and the SA error. Whenever the space covering was good, the error was low. On the other hand, poor coverage generated both high and low errors. This is mainly due to the estimator of coverage itself. Since the method of the Maximal Volume (described in section 3.4.1.5) is related to the smallest distance among all points, the quality of the space coverage may be underestimated if most of the points are well distributed but two or more are placed close to each other.

3.7. Conclusion

In this chapter it was discussed different strategies of sensitivity analysis. In the context of evaluating models for e-beam lithography processes, the strategy of choice is the variance-based sensitivity analysis. The choice relies on the structure of the models used, which present usually less than 20 inputs and their run time for this kind of evaluation is less than a second. Since most of the e-beam models are not suitable for the direct calculation of the sensitivity indices, two different approaches for estimating the SA were presented and compared. The approach proposed by Saltelli/Sobol' presents a faster convergence to the precise value than the scatterplot approach, being selected to be used in this thesis.

Chapter IV. Lithography Modeling

"The only relevant test of the validity of a hypothesis is comparison of prediction with experience"

- *Milton Friedman*

4. Lithography Modeling

E-beam lithography is widely used today by the industry for writing masks. Understanding the e-beam process characteristics and being able to predict the results of such process is a cornerstone for high resolution lithography both for today and for future technology nodes.

For this purpose, it is necessary to understand the most relevant characteristics of e-beam processes, identifying which ones require modeling and so on. When it comes to modeling exposure processes, there are two different strategies. The first one consists of using first principle models, which are based on the laws and theories related to electron-matter interactions, chemical behavior of the resist, acid diffusion, etc. These models are usually based on formulas describing each contributor factor and combined to Monte-Carlo simulations that address all probabilistic elements on the formulas. The second strategy uses empirical relations based on experimental observations and attempts to define a law "as simple as possible" that describes the behavior of the exposure. The models used in this strategy are usually known as compact models.

The characteristics of e-beam lithography are presented in the first section of this chapter. In the next two sections both first principle models and compact models are described and discussed. Finally, a comparison between these two types of models is also presented.

4.1. E-beam process characteristics

The electrons sent by e-beam machines interact with each other and with the matter. Figure 54 shows a representation of the most significant phenomena in a graphical way.

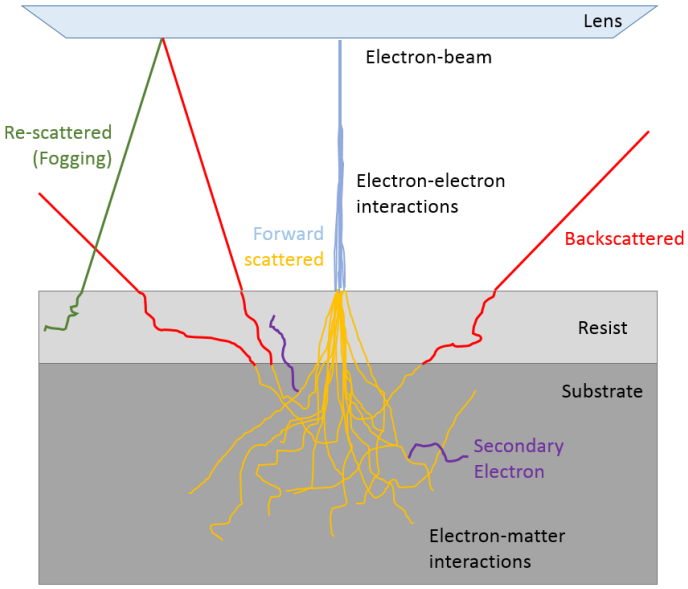


Figure 54. Representation of the forward scattering, backscattering and secondary electron generation by an electron-beam and its interactions.

4.1.1. Forward Scattering

While being transmitted from the electron gun to the material, the electrons present in a beam may suffer from lens aberrations and also experience coulomb interaction, which causes deflection by small angles but enough to statistically broaden the beam when it reaches the material. This phenomenon is known as electron forward scattering and its range may vary significantly, but is in the order of 2nm to 40nm, depending basically of the energy used to accelerate the electrons, of the beam current and of the lens aberrations. Figure 54 shows in gray the electron-electron iteration that causes the beam diameter to increase and in yellow the continued scattering of the electrons after reaching the material.

4.1.2. Backscattering

When the electrons finally reach the material, a second phenomenon is produced related to the electron-material iterations. Electrons passing close enough to an atom undergo attraction and scattering, changing their trajectory. After this interaction the electrons may continue with enough energy to travel inside the material and interact with other atoms on their way. In their course electrons may give energy to the resist up to several micrometers away from their incident point. This phenomenon is known as electron backscattering and is basically dependent of the energy used to extract the electrons from the source and the materials with which it interacts. Figure 54 shows in red the trajectories of examples of backscattered electrons.

4.1.3. Secondary Electrons

While passing through matter, an incident electron (primary electron) may hit another electron that was already present in the material (secondary electron), with enough energy to eject it from its location and provide it some energy to have a trajectory of its own inside the matter. The generated secondary electron may present enough energy to reach and give energy to the resist. Figure 54 shows in purple the trajectories of examples of secondary electrons generated by primary electrons.

4.1.4. Flare or Fogging

The phenomenon known as Fogging is the re-scattering of backscattered electrons back to the material. This is usually due to the reflection of such electrons on the pole piece of the objective lens of the electron-beam system. Those electrons return to the electron sensitive resist up to several millimeters away from their initial incidental point. Although very few electrons are re-scattered in a given region, their low energy makes their impact sufficient to still have a significant impact on the printed patterns dimensions since the energy loss by an electron is inversely proportional to its energy. Figure 54 shows in green an example of re-scattered electron.

4.1.5. Acid Diffusion on Chemically Amplified Resists (CARs)

In order to improve the writing time it is interesting to employ the most sensitive resist possible. In this way the amount of energy required to draw a pattern on the resist is smaller, saving exposure time. One strategy to improve the sensitivity of the resist is called chemical amplification. In this process, acid molecules are released during the exposure. Then, the post exposure bake process activates these acid molecules in a catalytic reaction that induces a solubility switch of the resist. Meanwhile, acids diffuse leading to a blur of the final aerial image. Since a single acid molecule can catalyze many of such reactions, lower dose is required, which means that less electrons need to be sent or, in other words, the beam need to spend less time exposing the same region [MACK, 1995].

The diffusion of the acid also improves line roughness and shot noise. On the other hand, the range of the diffusion is itself a limitation to the process resolution and work as a blur over the original exposure result [OKOROANYANWY, 2004].

4.1.6. Etching

Etching is a technique used to remove layers of a given material from the surface of a wafer or a mask. Etch processing is done by liquids (wet etching) or using etchant gases (dry etching). The first was more often used in reticle manufacturing while the second is more common in wafer manufacturing and advanced maskshops [GRANIK, 2001].

The wet etching approach relies on liquid-phase etchants where usually the wafer or reticle is immersed in. The dry etching approach consist of particles (either ions or radicals) colliding with the surface, assisting chemical reactions and detaching material from it. The process is performed until enough material is removed. Photoresists patterned on the surface may protect certain areas from this reaction. One key aspect of the etching process is the selectivity, which is the characteristic of impacting a given material without interacting with another. Ideally, the resist layer would not be impacted during etching, which would only remove the material from the target layer. This is usually not the case and the impact over the resist is one of the sources of resolution impact over the resulting pattern.

Another main aspect of the etching process is the horizontal and vertical etch rates, known as lateral etch ratio (R_L), which will determine the impact over the final resolution and may be computed as shown in (57).

$$R_L = \frac{Horizontal\ Etch\ Rate\ (r_H)}{Vertical\ Etch\ Rate\ (r_V)} \tag{57}$$

A R_L value equal to one identifies a case of isotropic etching where both horizontal and vertical directions are etched on the same rate. In this case, the bias caused by etch is more significant. A R_L value equal to zero identify a directional etch, which is ideal for resolution purposes but is not feasible in practice. What is feasible is to obtain an anisotropic etching, which presents a larger rate for vertical etch but still presents some resolution impact due to the horizontal etch

rate. A simplified representation of these three cases, considering ideal selectivity, is presented in Figure 55.

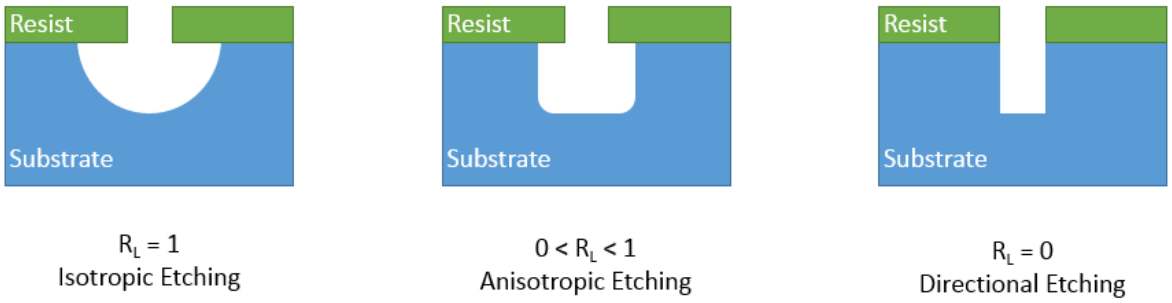


Figure 55. Pictorial representation of the impact of etching over the substrate for the isotropic, anisotropic and direction cases.

The etching behavior depends on the pattern CD, of the characteristics of the neighboring patterns, on the resist characteristics, among others.

4.2. First Principle Models

First principle models are implementations of the physical and chemical laws and theories related to the process steps. Usually, one or several models are associated to each step, which are not necessarily all simulated with the same level of precision.

The main steps that are frequently included in simulations of first principle model for lithography are the e-beam scattering (forward and backward) and the resist exposure and development. The photo-acid generation (PAG) and the post-exposure bake process (PEB) are increasing in importance as the model precision requirements grow. A flow of the most relevant steps to be simulated using first principle models is presented in Figure 56.

Figure 56. Sequence for simulating an e-beam lithography relying on First Principle Models.

4.2.1.1. Electron Scattering Monte Carlo Simulation

The forward and backward scattering may be simulated by a series of phenomena that may occur according to certain laws of probability. The common strategy to account for these statistical probabilities is to perform the simulation of the trajectories of thousands to millions of electrons, in a way that the beam behavior may be accurately modeled. The random behavior is usually computed by means of Monte Carlo simulation. Software such as CASINO [HOVINGTON, 1997a], [DROUIN, 2007] and CHARIOT [BABIN, 2003] are frequently used to describe the distribution of scattered electrons in a resist.

The trajectory of each electron is independently simulated, accounting for the different interactions with the different materials, the probability of generating secondary electrons, the amount of energy loss over each collision and so on. The sequence performed for computing a single electron trajectory is shown in Figure 57. Although the single trajectory flowchart is based in a version described by Hovington [HOVINGTON, 1997a], most of the examples of calculation in this section are from Kyser [KYSER, 1974].

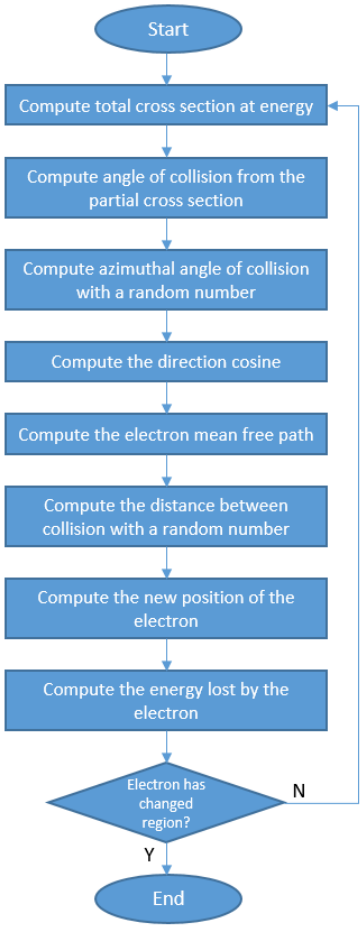


Figure 57. Flowchart diagram from the computation of the trajectory of a single electron in the Casino Software, from [HOVINGTON, 1997a].

The first step is the computation of the total cross section. A cross section is the effective area that indicates the probability of some scattering or absorption event to occur. The total cross section is the summing over the cross sections for different angles, given by the integration of the cross section equation for angles ranging from 0 to π. One of the most frequently used equations for computing a cross section is the one created by Rutherford, which is given by (58).

$$\sigma_i'(\theta) = \frac{d\sigma_i(\theta)}{d\Omega} = \frac{Z_i(Z_i+1)e^4}{4E^2(1-\cos\theta+2\beta_i)^2} \quad (58)$$

Where Z_i is the atomic number, e is the electron charge, E is the incident electron energy, θ is the scattering angle and β_i is the screening parameter to account for electrostatic screening of the nucleus by the orbital electrons.

In order to determine the angle for a particular scattering event, a random number is generated and used to check the probability of scattering by any given angle. The probability is given by (59).

$$P(\theta)d\Omega = \frac{\sigma_i'(\theta)d\Omega}{\sigma_i} \quad (59)$$

Where σ_i is the total cross section.

And, therefore, the angle is given by computing the directional cosine. Considering $F(\theta)$ the indefinite integral of $P(\theta)d\Omega$, the relation is given by (60).

$$cos(\theta) = 1 - \left[\frac{2\beta F(\theta)}{1+\beta - F(\theta)}\right] \quad (60)$$

The next step is to compute the mean free path between scattering events. It may be computed as shown in (61).

$$\Lambda = \frac{1}{n\sigma} = \frac{A}{N_A \rho \sigma} \quad (61)$$

Where n is the volume density of atoms, σ is the total cross section for scattering, N_A is the Avogadro's number, A is the atomic weight and ρ is the mass density.

The energy loss between scattering events may be computed using different approximations. One of the most frequently used is the continuous-slowing-down approximation of Bethe (62). It is an approximation since it assumes that instead of the slowing down be related to events that occur with the electron on its way, it is a continuous effect on the distance.

$$\frac{dE}{dS} = \frac{-2\pi e^4 \rho N_A}{E} \frac{1}{\sum}\left[\frac{C_i Z_i}{A_i} \ln\left(\frac{\gamma E}{J_i}\right)\right] \quad (62)$$

Where $\gamma = 1.166$ for relativistic energy electrons and J_i is the mean ionization energy.

The first iterations of the flow presented in Figure 57 is presented in Figure 58, from [KYSER, 1974].

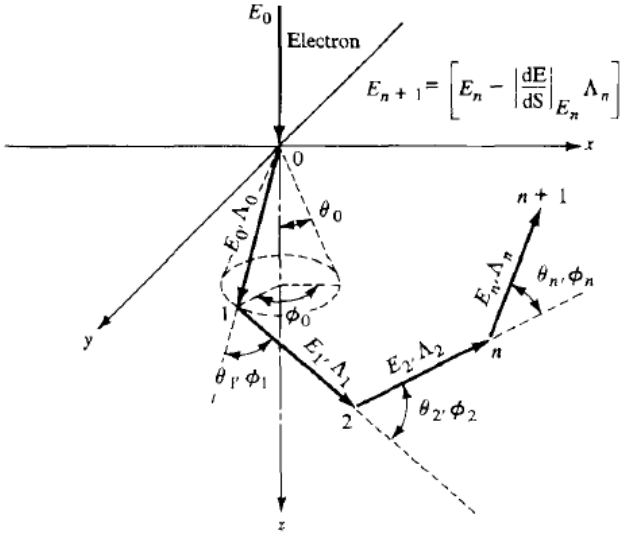


Figure 58. Geometry of the initial steps of electron scattering and energy loss, showing the values computed for generating a single electron trajectory, from [KYSER, 1974].

Usually the computation of the trajectory of an electron continues until the energy of the electron is below a given threshold. Each trajectory is computed independently and may be parallelized. An example of the computation of 200 electron trajectories is shown in Figure 59 from a run in Casino Software version 2.4.8.1 for a 200nm PMMA resist over a Silicon substrate and a beam energy of 50keV and beam radius of 20nm.

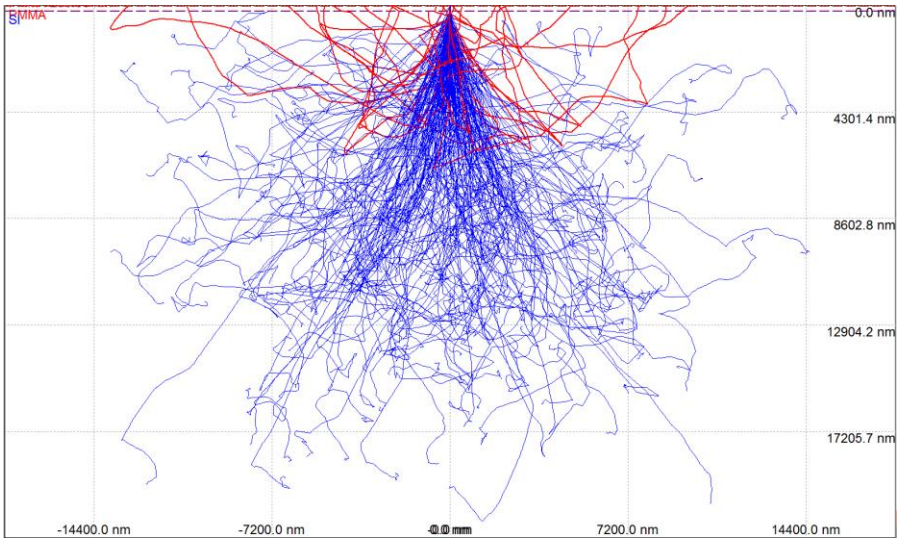


Figure 59. Trajectory of 200 electrons generated using Casino Software version 2.4.8.1.

4.2.2. Exposure Simulation

Exposure simulation accounts for describing the result of the exposure of a pattern over the resist. It may be performed by evaluating the electron scattering over an entire pattern or relying on the convolution of a Compact Model (PSF) that may be generated based on the results of the simulation of electron scattering, as described in section 4.2.1.1.

If the option is to rely on the Monte Carlo simulation, the procedure is quite similar to the one presented on the previous step, but instead of a single incident point, the electron beam is used to expose the entire pattern. The version of CASINO 3 [DEMERS, 2011] is able to represent the electron scattering in 3D, describing complex shapes and stacks, which is particularly useful for direct write applications were previously exposed layers impact the exposure results.

4.2.3. PAG Simulation

The next step is the photo-acid generation modeling. The Dill model [LIU, 2007] describes the kinetics of this phenomenon, which may also be described by more specific ones on the case of Chemical Amplified Resists [KRASNOPEROVA, 1994], [CROFFIE, 2000].

As long as no in or out diffusion occurs, the concentrations of photo-acid generator (PAG) and acid (A) are directly related to each other, and one may describe as a normalized relation as (63) [ERDMANN, 2001].

$$[\mathrm{PAG}] + [\mathrm{A}] = 1 \tag{63}$$

In the case of a positive resist, one may describe the PAG concentration by the equation presented in (64) [SCHNATTINGER, 2006].

$$[\mathrm{PAG}] = e^{-C_{Dill}D} \tag{64}$$

Where C_{Dill} is the resist sensitivity and D is the dose.

4.2.4. PEB Simulation

Amplification reaction occurs during the baking procedure. In positive tone resists, for instance, as an effect of the temperature, the acid diffuses and continues removing the protection of areas that were not initially exposed. The final result is that the exposed patterns become larger than the original exposure. The diffusion is a function of the time of PEB, which must be considered during the simulation.

The PEB model for positive tone resists is based in a system of three differential equations [TOLLKÜHN, 2006], as shown in equations (65), (66) and (67).

$$\frac{\partial [M]}{\partial t} = -k_1 . [M]^p . [A]^q - k_2 . [M] \tag{65}$$

$$\frac{\partial [A]}{\partial t} = -k_3 . [A]^r - k_4 . [A] . [B] \tag{66}$$

$$\frac{\partial [B]}{\partial t} = -k_4 . [A] . [B] \tag{67}$$

Where A is the acid, B is the quencher base, M is the dissolution inhibitor, k_1 is the rate constant for acid catalyzed site deprotection, k_2 is the rate constant accounting for thermally induced deprotection, k_3 is the acid loss reaction rate, k_4 is the acid neutralization rate, exponents p and q are reaction order of each species while exponent r is an effective reaction order.

Besides from the deprotection kinetics, the heating of the resist also causes diffusion of the acid and base components inside the resist. The diffusion of acid, or of any other resist component, may be described by (68).

$$\frac{\partial [A]}{\partial t} = -\nabla (D([A])\nabla [A]) \tag{68}$$

Where D specifies the diffusion coefficient.

This system covers most of the reaction kinetic models for CAR. However, this model is quite complex and one may implement a simplified version of it [MÜLDERS, 2005].

The computation of this kind of model is often expensive because it passes by discretizing the terms as a function of time and then performing the calculation iteratively for each interval Δt.

4.2.5. Resist Development Simulation

The solubility of the resist is characterized by a development rate as a function of the inhibitor concentration. This rate express the speed by which the development will occur and it is inversely proportional to the concentration of inhibitor in the region. One of the first approaches was presented by Dill [DILL, 1975], which consisted of an empirical curve fit of experimental development rate data. There are several different models to compute this rate [ARTHUR, 1997][ARTHUR, 1998][HENKE, 1999][KARAFYLLIDIS, 2000].

A simplified model is the one known as Notch-model shown in (69) [MACK, 1998]. It was reported to provide the best fit over experimental data from development rate monitor (DRM) [ERDMANN, 2001].

$$R([M]) = \mathrm{R}_{max}(1-[\mathrm{M}])^{n}.\left[\frac{(a+1).(1-[M])^{n\_notch}}{a+(1-[M])^{n\_notch}}\right] + \mathrm{R}_{min}\ ) \tag{69}$$

Where R_{min} and R_{max} are the minimum and maximum development rate of the resist, n is the steepness of the rate curve, n-notch is a measure of the strength of the notch and a is given by (70).

$$a = \frac{n\_notch + 1}{n\_notch - 1}\left(1 - \mathrm{M}_{th\_notch}\right)^{n\_notch} \tag{70}$$

Where M_{th_notch} is the value of [M] where the notch.

This type of model requires high computational effort since it is evaluated in several iterations as fractions of time in the process. The smaller each fraction of time is, more precise is the simulation but more iterations are required to complete the simulation time.

4.3. Compact Models

Another strategy of modeling lithography processes relies on Compact Models. Compact Models are usually a faster alternative to First Principle Models and, if properly conceived and calibrated, may provide useful and accurate results.

4.3.1. Electron Scattering

Electron scattering is usually modeled using a Point Spread Function (PSF). It describes how the energy of the electrons spreads out in the resist and the substrate. In other words, a PSF represents a dose distribution in the resist. This dose distribution is related to the characteristics of the lithographic process and to the repartition of the electrons in the material which is a consequence of the electron scattering events. It is possible to summarize a PSF as a sum of two contributions, one describing the forward scattering phenomenon, associated with short range effects, and the other describing the backscattering phenomenon, associated with long range effects, as shown in (71).

$$PSF(x) = ShortRange(x) + LongRange(x) \tag{71}$$

Observe that the resulting PSF is a normalized distribution, assuring that the total energy represented is equal to one. The PSF is then convoluted to the patterns in order to generate a dose profile in the resist. Figure 60 shows an illustration of a convolution of a pattern and a PSF in order to compute the resulting energy distribution after exposure.

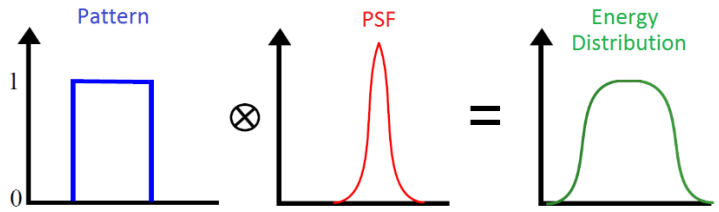


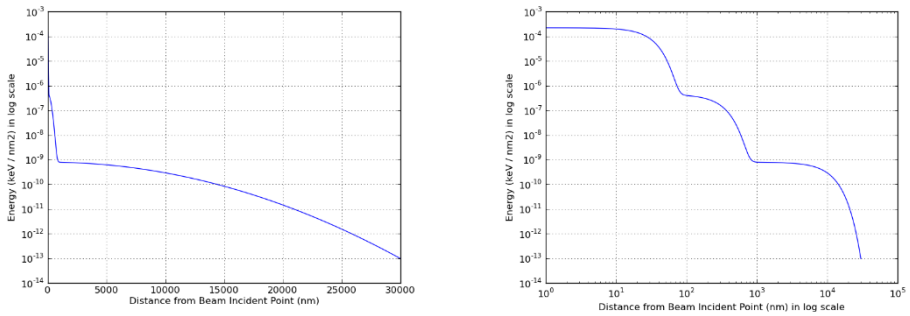
Figure 60. An illustration of the convolution of a pattern and a PSF model to generate an energy distribution [MARTIN, 2011].

4.3.1.1. Gaussian PSF

One of the most frequently employed PSF is composed of 2 Gaussian distributions, one used to represent the forward scattering while the other explaining the backscattering effect [CHANG, 1975][KYSER, 1975]. Eventually the backscattering effect became more complex (either by occurring in different ranges or by presenting a behavior difficult to be explained by a single Gaussian distribution) and more Gaussian distributions were added in order to improve the model precision. The addition of one or more Gaussian distributions in the backscattering portion of the PSF helps in retrieving an acceptable representation of the experimental data. The Sum-of-Gaussians PSF model, which is currently the state-of-the-art of the electron lithography community, is presented in equation (72).

$$PSF(x) = \frac{1}{1+\sum \eta_i}\left[\frac{1}{\pi\alpha^2}\exp(-x^2/\alpha^2) + \sum \frac{\eta_i}{\pi\beta_i^2}\exp(-x^2/\beta_i^2)\right] \quad (72)$$

An example of the 3 Gaussian PSF model is presented in Figure 61.



(a) (b)

Figure 61. An example of a 3 Gaussian PSF model (a) in a semi-logarithmic scale and (b) in a logarithmic scale in both axes. The parameters are α = 30nm, β1 = 300 nm, η1 = 0.2, β2 = 10000 nm and η2 = 0.4

The resulting PSF is a very simple yet powerful way to describe the electron scattering, presenting a performance several orders of magnitude faster than computing all the operations described to model the electron scattering using First Principle models. However, for more recent technology nodes, this model is no longer accurate enough and different approaches were required to compromise between precision and performance [KAMIKUBO, 2010][TANABE, 2010] [FIGUEIRO, 2012a].

4.3.1.2. Gamma PSF

Although the Gaussian PSF has been used for years, it is known that is lacks some precision when it comes to predict the effects of electron scattering [FIGUEIRO, 2012a]. In order to illustrate this, a First Principle model is used as reference. It is the CASINO software described in previous section that will be used to observe the energy distribution on the resist. Two different stacks are proposed for this test, as shown in Figure 62, one representing the Chrome On Glass (COG) Stack corresponds to the stack of a traditional photomask, comprising a Chromium layer over a glass substrate. The other is called the EUV Stack, which describes in a simplified way the stack of an Extreme Ultraviolet mask (EUV), consisting of a resist layer, a heavy metal material (Ta), on a SiO2 substrate. All other inputs of the CASINO software were identical, namely number of simulated electrons (5 millions), acceleration voltage, (50keV), beam radius (20nm), and tilt of the sample (0^o).

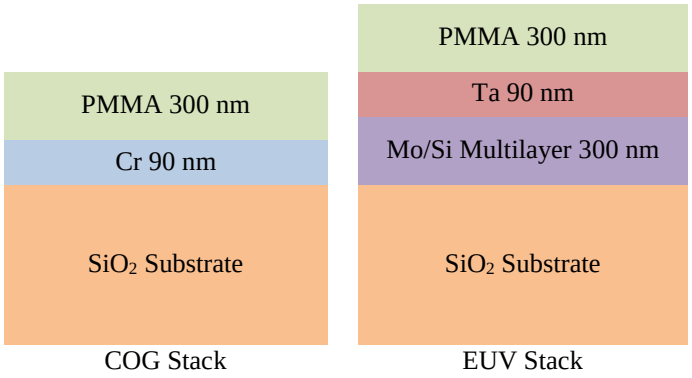
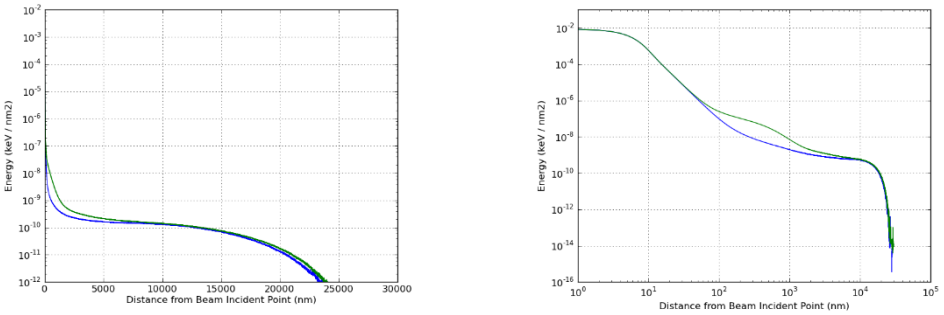


Figure 62. Different stacks used for simulating the surface radius of backscattering electrons.

The PSF data shown in Figure 63 correspond to COG and EUV stacks. Both curves overlap perfectly except for mid-range positions. The forward scattering, occurring nearby the beam (radius < 100nm), is indeed mainly dependent on the resist material and thickness. The long-range backscattering (radius > 10µm) is also independent from the film structure, because electrons, in this case, are backscattered from the substrate. The difference between both curves would then be caused by the mid-range backscattered electrons [KAMIKUBO, 2010][TANABE, 2010]. Therefore, in order to obtain more accurate representations of this phenomena it is necessary to provide a model able to describe it using the smallest number of parameters possible.



(a) (b)

Figure 63. PSF Model - (a) in linear scale in the x axis and in logarithmic scale in the y axis and (b) in logarithmic scale in both axes - for COG stack in blue and EUV stack in green. In the detail the main difference between both curves.

Observe by comparing the Gaussian PSF (in Figure 61) and the actual energy distribution (in Figure 63) that there is a significant difference on the way the mid-range behaves. The Gaussian PSF, when plotted in a log-log scale presents a behavior similar to a staircase (Figure 61 (b)). It does not present the flexibility to generate a heavy tail as observed in Figure 63 (b) and, moreover, is not capable of generating local peaks, as seen in the EUV stack (green plot in Figure 63 (b)). For this reason, a different PSF model is proposed in this thesis based on the Gamma distribution [SCHIAVONE, 2012].

The Gamma distribution is a probability distribution described by two parameters, the scale parameter (θ) and the shape parameter (κ). The graphs presented in Figure 64 shows several Gamma distributions for different κ and θ values.

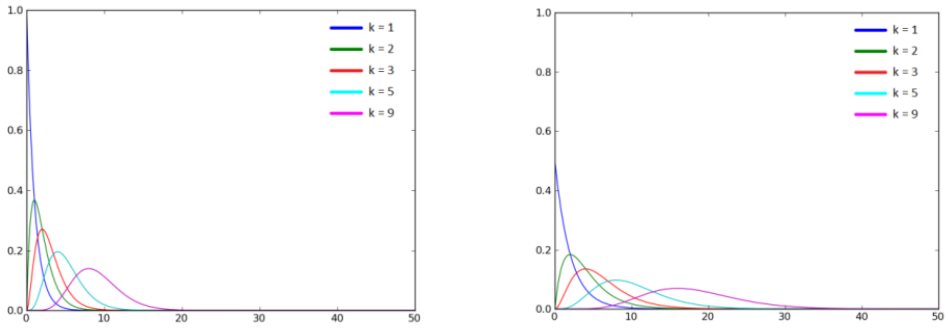
The probability distribution function of a Gamma distribution is given by (73).

$$PDF(x) = x^{\kappa-1}\frac{exp(^{-x}/_{\theta})}{\Gamma(\kappa)\theta^{\kappa}} \tag{73}$$

Where Γ is the gamma function, represented as shown in (74).

$$\Gamma(x) = \frac{e^{-\gamma x}}{x}\prod_{k=1}^{\infty}\left(1+\frac{x}{k}\right)^{-1}e^{x/k}, \qquad \gamma \approx 0.577216 \tag{74}$$

In Figure 64, from the off-centered characteristic and the versatility of the Gamma function, it is possible to observe its potential to describe the backscattering phenomena, especially the off-centered peaks observed in some of the Monte Carlo simulations. Therefore, the next step is to include them in a PSF description.



(a) (b)

Figure 64. Different Gamma distributions for κ values [1, 2, 3, 5, 9] in every chart and θ values of (a) θ = 1 and (b) θ = 2

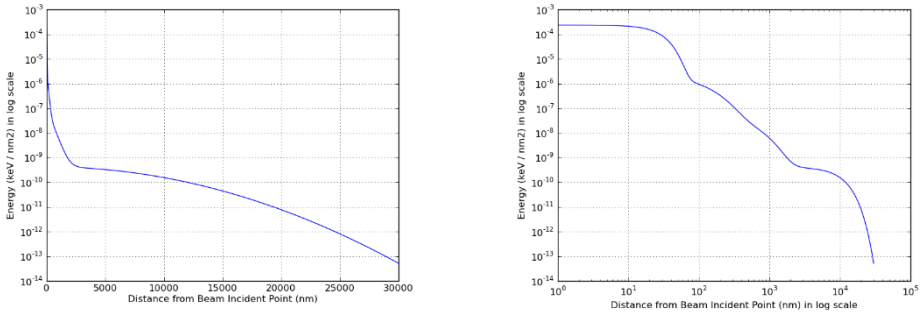
Some of the properties of the Gamma PDF is that it's standard deviation is given by $\kappa\theta^2$ while the position of the peak (the mode of the distribution) is given by $(\kappa-1)*\theta$. These two characteristics show how connected these two values are, which may difficult the calibration.

As briefly discussed earlier, an electron PSF may be seen as a sum of components representing different physical phenomena associated to different ranges. In the sum of Gaussians PSF, for instance, one may say that the first Gaussian is related to the short range effect, while the last is related to the long range. If there are more than two Gaussians, the others are usually employed to explain what happens in between those two distributions, or, in the mid-range. As it was possible to observe in Figure 63, the difference between the EUV and COG mask stacks simulation were significant only in the mid-range.

Therefore, by replacing the Gaussians that are inserted between the first one (s*hort range* portion) and the very last one (*long range*) by a sum of Gamma distributions, properly normalized, the resulting PSF is given by (75) [FIGUEIRO, 2012].

$$\mathbf{PSF\_gm(x)} = \frac{1}{1+\sum \eta_i + \eta_f}\left[\frac{1}{\pi\alpha^2}\exp\left(-x^2/\alpha^2\right) + \sum \frac{\eta_i}{2\pi\kappa_i\theta_i}\, x^{\kappa_i - 1}\frac{exp\left(-x/\theta_i\right)}{\Gamma(\kappa_i)\theta_i^{\kappa_i}} + \frac{\eta_f}{\pi\beta_f^2}\exp\left(-x^2/\beta_f{}^2\right)\right] \quad (75)$$

An example of the PSF Model energy distribution result is presented in Figure 65.



(a) (b)

Figure 65. An example of a 2 Gaussian 2 Gamma PSF Model (a) in linear scale in the x axis and in logarithmic scale in the y axis and (b) in logarithmic scale in both axes. The parameters were $\alpha = 30$, $\kappa_1 = 1.2$, $\theta_1 = 80$, $\eta_1 = 0.2$, $\kappa_2 = 1.5$, $\theta_2 = 300$, $\eta_2 = 0.1$, $\eta_3 = 0.2$ and $\beta_3 = 10000$.

4.3.1.3. Voigt/Pearson PSF

Another approach considered in this work is to describe the energy distribution using a Voigt distribution [TORTAI, 2012]. The Voigt distribution is a probability distribution which is composed by the combination of two other distribution profiles, a Gaussian distribution and a Lorentzian distribution. The Voigt profile is the convolution of these two profiles, being able to assume a shape very similar to one or the other, according to the parameters used. This flexibility makes the Voigt distribution an interesting candidate to implement a PSF model. One description of the Voigt function is given in (76).

$$\text{Voigt}(x, \sigma, \gamma) = \int_{-\infty}^{\infty} Gauss(x', \sigma) Lorentz(x - x', \gamma) dx' \quad (76)$$

Where $\text{Gauss}(x', \sigma)$ is the centered Gaussian profile with standard deviation σ given by (77).

$$Gauss(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-x^2/2\sigma^2\right) \quad (77)$$

And $\text{Lorentz}(x - x', \gamma)$ is the centered Lorentzian profile with standard deviation γ given by (78).

$$Lorentz(x, \gamma) = \frac{\gamma}{\pi(x^2 + \gamma^2)} \quad (78)$$

The Voigt profile may be computed as shown in (79)

$$Voigt(x, \sigma, \gamma) = \frac{\mathbf{Re}[\mathbf{w}(\mathbf{z})]}{\boldsymbol{\sigma}\sqrt{\mathbf{2}\boldsymbol{\pi}}} \tag{79}$$

Where Re[w(z)] is the real part of the Faddeeva function [Faddeeva, 1954] evaluated for $z = (x + i\,\gamma)/\,\sigma\sqrt{2}$.

This Voigt formulation itself is not much suitable for implementing a model for computing proximity effect correction because it is computational expensive [TORTAI, 2012]. However, the Pearson type VII distribution is a close approximation of the Voigt profile. It is given by (80).

$$PearsonVII(x, \sigma, \nu) = \frac{\mathbf{1}}{\boldsymbol{\sigma}\sqrt{\mathbf{2}\boldsymbol{\nu}-\mathbf{3}}\boldsymbol{B}(\boldsymbol{\nu}-\mathbf{0.5},\mathbf{0.5})}\left[\mathbf{1} + \left(\frac{\boldsymbol{x}^{\mathbf{2}}}{\boldsymbol{\sigma}\sqrt{\mathbf{2}\boldsymbol{\nu}-\mathbf{3}}}\right)^{\mathbf{2}}\right]^{-\boldsymbol{\nu}} \tag{80}$$

Where B is the Beta function, σ is the standard deviation and ν is the parameter that controls the kurtosis of the distribution. If ν approaches infinity, this function actually approaches the Gaussian distribution.

After computing the normalization of this Pearson type VII distribution, one may obtain a PSF based only on this function, as presented in (81).

$$PSF\_p\,(x) = \frac{\mathbf{1}}{\sum \boldsymbol{\eta}_{\boldsymbol{i}}}\left[\sum \frac{(\boldsymbol{\nu}_{\boldsymbol{i}}-\mathbf{1})}{(\boldsymbol{\sigma}_{\boldsymbol{i}}\sqrt{\mathbf{2}\boldsymbol{\nu}_{\boldsymbol{i}}-\mathbf{3}})^{\mathbf{2}}\boldsymbol{\pi}} * \left(\mathbf{1} + \frac{\boldsymbol{x}^{\mathbf{2}}}{(\boldsymbol{\sigma}_{\boldsymbol{i}}\sqrt{\mathbf{2}\boldsymbol{\nu}_{\boldsymbol{i}}-\mathbf{3}})^{\mathbf{2}}}\right)^{-\boldsymbol{\nu}_{\boldsymbol{i}}}\right] \tag{81}$$

Observe that although all components of this distribution were described as Pearson type VII distributions, but if a Gaussian distribution is enough to explain the behavior of the electron scattering, the ν parameter may be set in a way that the component becomes similar to a Gaussian distribution. An example of a 4 Pearson PSF is shown in Figure 66 (in green), while fitting an energy distribution from CASINO (in black). The different Pearson distributions are also represented in magenta.

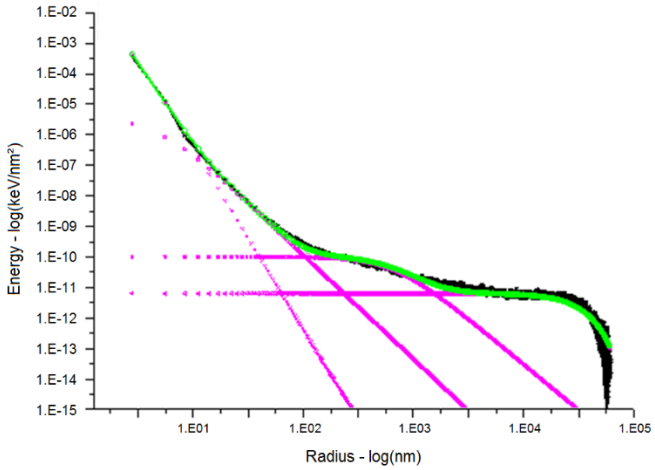


Figure 66. Example of a 4 Pearson PSF (in green) calibrated based on an energy distribution from CASINO simulation (in black). The four distributions are separately shown in magenta.

4.3.1.4. Spline PSF

Another strategy to represent the energy distributed in the resist proposed in this thesis work is to use smoothing cubic splines [JEDIDI, 2013]. Smoothing cubic splines has been introduced during the sixties [REINSCH, 1967]. A cubic spline function with knots at x_1, x_2,…, x_n where the knots are ordered inside a finite interval [a, b], $a = x_1 < x_2 < x_n = b$, is a function g of a class $C^2[a, b]$ that is equal to a third order polynomial g_k on each subinterval $[x_i, x_{i+1}]$. In other words, it is a piecewise function that satisfies the following properties:

$$\forall\, i \in [0, n-1],\ \ \exists\, (a_i, b_i, c_i, d_i),\ \ \forall\, x \in [x_i, x_{i+1}],$$

$$g(x) = g_i(x) = \ a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

Where g, g' and g'' are continuous on $[a, b]$, implying that at each knot $(x_i)_{1 \le i \le n-1}$,

$$g_{i-1}(x_i) = \ g_i(x_i)$$

$$g'_{i-1}(x_i) = g'_i(x_i)$$

$$g''_{i-1}(x_i) = g''_i(x_i)$$

The latter properties lead to a linear system of $(4n - 2)$ independent equations for $(4n)$ unknown coefficients. To make the system full rank and g uniquely defined, 2 extra-equations are needed. They are provided by the boundary conditions, which hold at the end points x_0 and x_n. The cubic spline is said to be "natural" if $g''(x_0) = g''(x_n) = 0$. The function g would have then no curvature at the extremities of the interval $[a, b]$. One example of spline with 5 knots is shown in Figure 67.

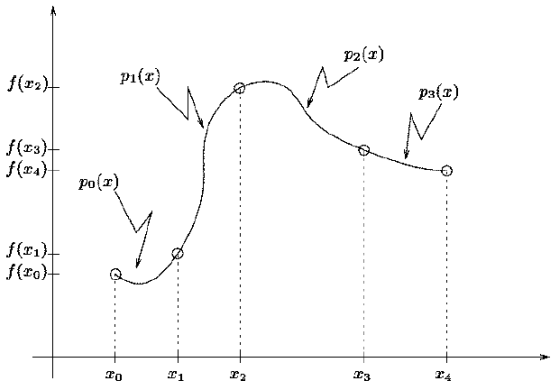


Figure 67. Example of a spline with 5 knots.

One example of a Spline PSF with 6 knots is shown in Figure 68.

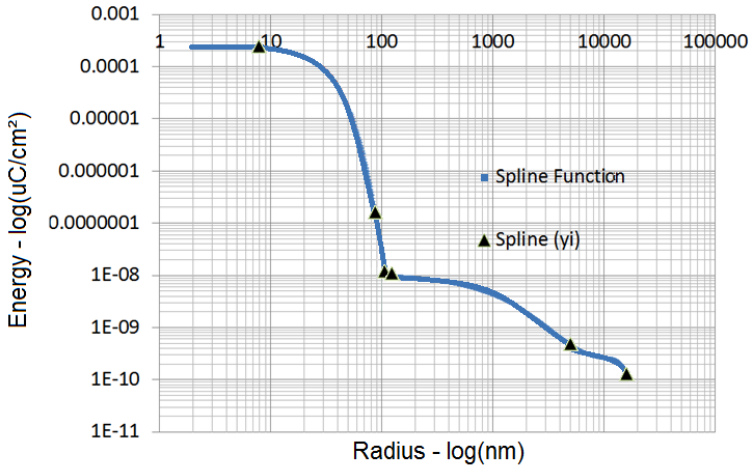


Figure 68. Example of a Spline PSF with 6 knots.

4.3.1.5. PSF Model Comparison

Among the strategies presented in this chapter, each presents its strong and weak points. The Gaussian PSF is still a very simple PSF that is able to explain a large portion of the phenomena with few parameters. For instance, a 2 Gaussian PSF is represented with only 3 parameters and 2 more parameters for each additional Gaussian distribution added. Gamma and Pearson PSFs present 3 parameters per distribution, except the first with only 2. For this reason, combining Gaussian and Gamma or Pearson distributions in a PSF is advantageous, using Gaussian distributions in any region where it is sufficient and more complex distributions everywhere else. Finally, the Spline PSF is more complex to compare with other approaches since it usually presents more parameters but add much more flexibility and does not rely on any a priori assumption about the shape of the PSF. However, in terms of calibration effort, both the Gamma PSF and the Spline PSF are difficult to handle due to the correlation of their parameters with different characteristics of the distribution profile. On the other hand, the Pearson PSF presents significantly independent parameters. Besides, the Pearson distribution is also able to present the

same behavior as a Gaussian distribution, allowing a first calibration of a model containing only Pearson distributions to be used to detect which portions may be simplified by replacing by a Gaussian distribution and which parts do not.

4.3.2. Extreme-long-range Effects

Besides the effects on short-range, mid-range and long-range described in the previous section, there are also effects presenting a larger range. For instance, some electrons after backscattering from the sample may re-scatter into it after reflecting on the electron lens, several millimeters away from the initial incident point. The energy which is deposited by these electrons back into the resist also impact the exposure quality. This is called fogging. Other effect that may present extreme-long-range effects is the Chemical-Mechanical-Polishing (CMP), which is also dependent of the density of the patterns. Yet another effect is the etch loading, which may range from millimeters to centimeters.

In the early stages of the e-beam lithography, extreme-long-range effects were not included in models, being just considered as a background dose and fixed on the selection of the exposure dose. However, as the requirements of precision increased, some strategies were proposed by the industry. The most common one is to include the fogging effect into the PSF that explains the electron scattering phenomena [HUDEK, 2007], [ABE, 2007], [SUNDBERG, 2010]. In this case, another Gaussian distribution is added with a sigma corresponding to the range of the fogging. An example of such strategy is shown in Figure 69.

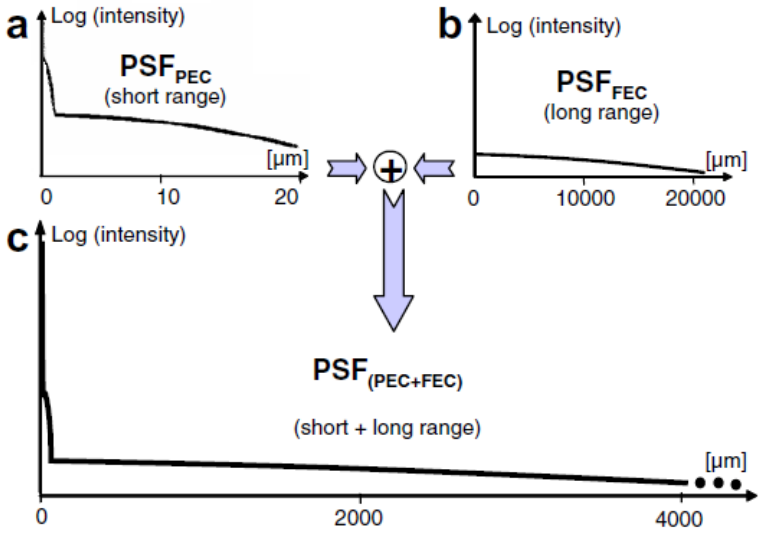


Figure 69. Example of the fogging model as a PSF_{FEC} joint with a PSF_{PEC}, which accounts for the electron scattering [HUDEK, 2007]

This approach is interesting because it uses the same structure of the PSF model, not requiring further implementation. However, the computation cost to compute the convolution of all shapes in the range of several thousands of microns may be prohibitive and some approaches based on density calculations have been proposed to improve the performance of such models [TEJNIL, 2008][FIGUEIRO, 2013b].

4.3.3. Etching Model

Besides describing how the energy profile in the resist depending on the energy distribution and the patterns exposed, another relevant aspect is the etching effect over the resist. Usually, etch loading effect is translated into a bias between the dimension observed after exposure and the one observed after etching. Simple models consider this bias constant over the entire circuit. As the technology advanced, some simple rules were used to describe the etch bias as a function of some pattern characteristics. For instance, it was observed a relation between the drawn CD and the etch bias, as shown in Figure 70 [STOBERT, 2013].

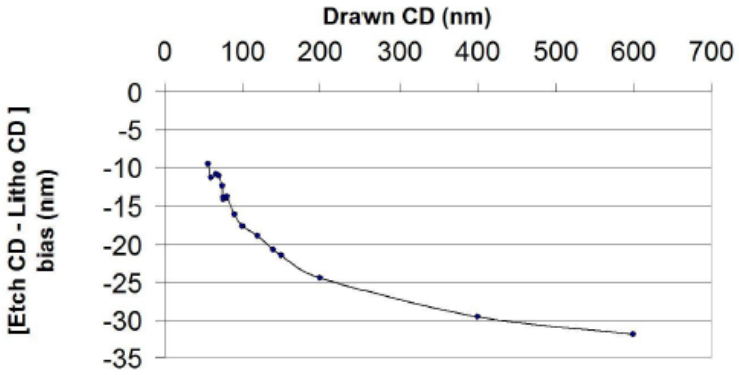
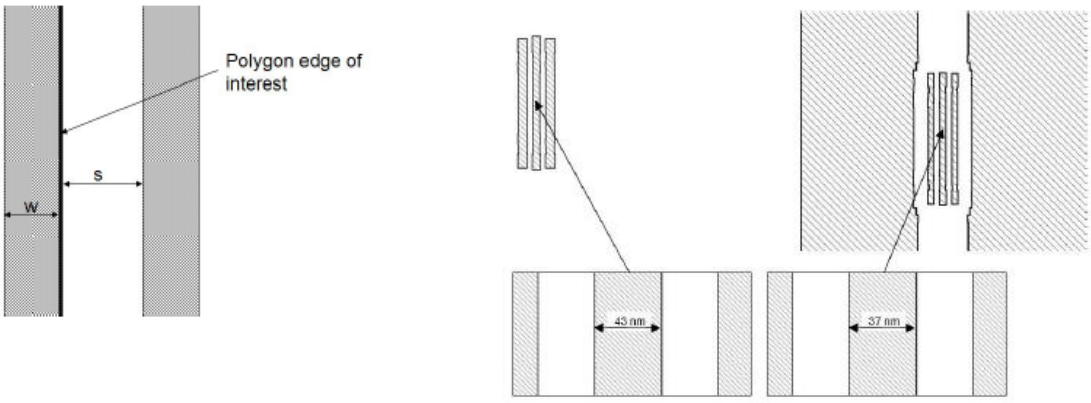


Figure 70. Example of influence of the CD of the drawn feature and the etch bias [STOBERT, 2013].

Besides the width of the pattern being etched, the space from the edge under evaluation and the closest feature also present some impact, as well as the local density of patterns. An illustration of these characteristics is shown in Figure 71.



(a) (b)

Figure 71. (a) Example of width and space characteristics that present impact over the final etch bias. (b) Different densities also cause different bias [STOBERT, 2013].

These characteristics were used on rule based bias descriptions but as technology advances this approach is becoming overly complex and employing a model is the alternative to describe complex conditions that impact etch bias [STOBERT, 2013]. The metrics such as the width (CD) and space are not trivial to be obtained for more complex pattern. Therefore, usually these

metrics are obtained by using different kernels that can be variants of Gaussian distributions computed as shown in Figure 72.

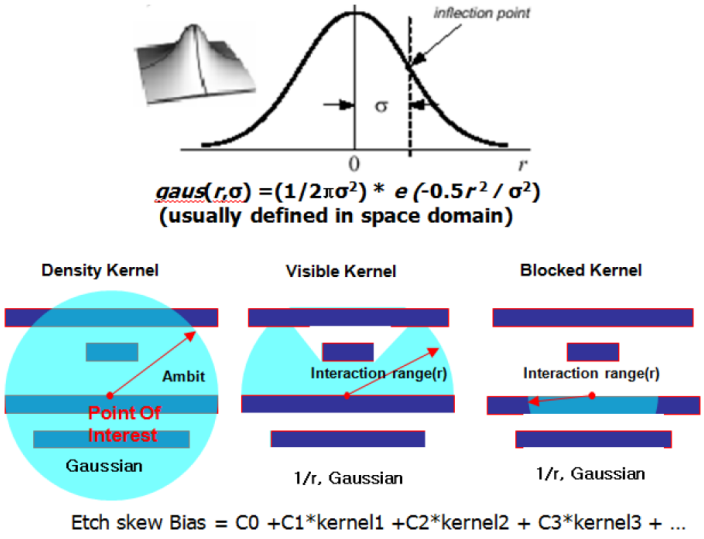


Figure 72. Example of kernel calculation for etch bias polynomial models [PARK, 2011].

Typically, three distinct kernels are used [PARK, 2011]. The density kernel accounts for etch loading effect caused by average pattern density within a given range from the point of interest (POI). The visible kernel represent etch loading effect caused by the open area near the POI. Finally, the blocked kernel accounts for the effect caused by the polygon width within interaction range near the point of interest.

There are several etch loading models available in the literature [DUNN, 2009][QINGWEI, 2010][PARK, 2011]. These models usually consist in computing a bias for each point based on polynomial functions which are a function of kernels that represent characteristics of the edge of interest. One example from [PARK, 2011] is shown in (82).

$$\text{Etch bias} = C0 + C1*\text{visible kernel} + C2*\text{density kernel} + C3*\text{blocked kernel} + \ldots\ldots \quad (82)$$

Using such model one may determine which bias should be applied to every edge on the pattern being simulated in order to predict the result after the etching procedure.

4.3.4. Resist Model

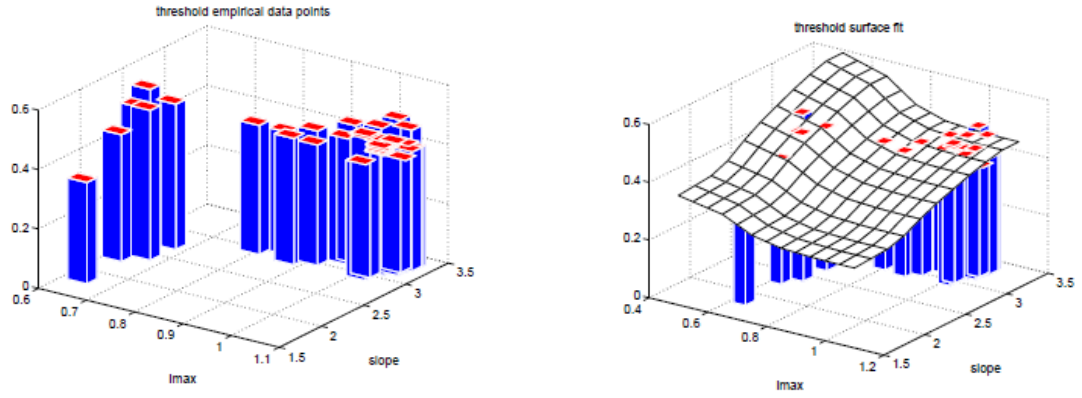
Besides describing how the energy may be distributed in the resist depending on the energy distribution and the patterns exposed, it is necessary to determine the final resist image, which is actually the resulting patterns developed in the resist.

Most of the compact resist models consist of a simple threshold, determining that energy values above such threshold are developed and those below are not (for a positive tone resist). Moreover, early models actually consider this as a constant threshold whose value is 0.5 in a normalized dose context when the PSF model has a total energy of 1.0.

A straightforward evolution of such model was the implementation of a constant threshold but not fixed in 0.5 but assuming a value resulting from the calibration procedure. This increases the versatility of the model and, even though it increasing the number of parameters to be calibrated, it has no significant impact on simulation time.

This approach reaches its limits for more advanced processes when there are significant differences on the resist behavior according to design characteristics. Variable threshold was proposed to continue to improve the precision of the resist models.

The first approach, known as Variable Threshold Resist Model [COBB, 1997] introduces the idea of resist threshold that is a function of the maximum local intensity (Imax) and the image slope. The model then relies on empirical data to determine the threshold value that would make the simulation meet the experimental value from line width measurements.



(a) (b)

Figure 73 (a) Empirical threshold values determined from experiments and (b) the surface fit described by VTR model superimposed on empirical data [COBB, 1998].

The main assumption on this case is that the threshold may be expressed in a polynomial form such as (83).

$$T = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(a_{i,j} . {I_{max}}^{i} \ {S_{max}}^{j} \right) \quad (83)$$

Where I_{max} is the peak value of the intensity cutline and S_{max} is the maximum slope of the intensity cutline. The parameters $a_{i,j}$ are calibrated based on the empirical data [COBB, 1998].

The polynomial expression contained, in some cases, more than 20 terms [RANDALL, 1999]. This approach was later extended by adding other parameters. For instance the model VTR Enhanced (VTRE) [SAHOURIA, 2000] and the Threshold-5 (VT5) [GRANIK, 2002]. Moreover, a variable threshold approach may be extended in order to include etch loading effect. For this purpose, one may translate the given bias into a variation on the threshold. This approach allows completely absorbing the etching model into the variable threshold resist model [GRANIK, 2003].

All this model variations rely on polynomial function which include as parameters information such as the maximum intensity, among others. The drawback of this approach is the computational effort to obtain such values for each point in the entire layout. For this reason a different strategy is proposed in this work in order to provide a variable threshold resist model with improved performance.

A specialized strategy of kernel was developed by us in order to achieve the required performance while identifying the circuit characteristics that are relevant to describe different behavior of resist image, etching or any other phenomena [SAIB, 2014c]. The approach consists of using a new type of kernel computed by convolution called oriented kernel. It relies on a function of angular deformation f(θ), which can be expressed by its Fourier series as shown in (84).

$$f(\theta) = \sum_{n=0}^{\infty} (a_n.\cos(n\theta) + b_n.\sin(n\theta)\,) \tag{84}$$

Where n is the order of the Fourier series, θ is the direction on the kernel space and a_n and b_n the Fourier coefficients calculated as shown in (85) and (86).

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(\theta).\cos(n\theta)\,d\theta \tag{85}$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} f(\theta).\sin(n\theta)\,d\theta \tag{86}$$

It is possible then to introduce an angle for a phase shift in order to adjust the direction of the kernel to an angle θ =π/2+ φ. This generates the expression shown in (87).

$$f(\theta + \varphi) = \sum_{n=0}^{\infty} \big(a_n.\cos\big(n(\theta + \varphi)\big) + b_n.\sin\big(n(\theta + \varphi)\big)\,\big) \tag{87}$$

In this condition, the oriented kernel is described as (88).

$$V = \int_0^{\infty}\int_0^{2\pi} K(r).f(\theta + \varphi).S(r,\theta)\,.dr\,d\theta \tag{88}$$

Where K(r) is an isotropic kernel (such as a Gaussian kernel), S(r, θ) represent the patterns around the region under evaluation returning a null value when outside of patterns and non-null when inside a pattern.

The computation based on Fourier series allow the computation of the oriented kernel V independently of the angle φ. This is shown in (89).

$$V = \sum_{n=0}^{\infty} (a_n . \cos(n\varphi) + b_n . \sin(n\varphi)\ ) \int_0^{\infty} \int_0^{2\pi} K(r) . cos(n\theta) . \mathrm{S}(r,\theta)\ . dr\ d\theta$$
$$+ [(-a_n . \sin(n\varphi) + b_n . \sin(n\varphi))] \int_0^{\infty} \int_0^{2\pi} K(r) . sin(n\theta) . \mathrm{S}(r,\theta)\ . dr\ d\theta \quad (89)$$

Allied with this approach it was developed a variable threshold surface which relies on these kernels or any other calculation in order to provide a variable threshold adapted to explain complex process behaviors with a fast compact model [SAIB, 2014b].

The procedure to build such model consists in two main steps. One that is to determine which parameters are relevant to describe the process. This may rely on input variables such as the ones used in etching models or directly in kernels, such as the ones described earlier in this section. By using the oriented kernels one may obtain better performance and continuous values, which are always preferable when performing such modeling. Although there is no restriction on how many input variables one may use, as of today it was not required more than three different metrics (CD Kernel, Space Kernel and Density Kernel) to achieve high precision modeling. The CD kernel computes the area inside the oriented kernel directed to the interior of the pattern, the Space kernel computes the area outside the oriented kernel while the density kernel is computed by an isotropic kernel.

The second step is to determine the threshold desired to each combination of the metrics. Usually a calibration is performed based on experimental data coming from exposure and measurement of patterns presenting the desired metrics followed by the calibration of a PSF model and, finally, determining which threshold value would provide the same simulation value observed on the experiment. The final result is shown in Table 11.

Table 11. Example of metric values and the correspondent threshold value to meet the experimental value.

Metric I	Metric II	Threshold Value
0.1	0.2	0.62
0.2	0.4	0.49
0.3	0.3	0.54
0.4	0.6	0.68
0.5	0.7	0.42
0.6	0.5	0.38
0.7	0.9	0.35
0.8	0.1	0.32
0.9	0.8	0.34

From this data coming from the evaluation of the metrics and the experimental observation to determine the threshold for each input combination one may perform an interpolation to fill the space as a surface. The interpolation algorithm used in this case was Kriging. The result after interpolation may be represented the graph shown in Figure 74.

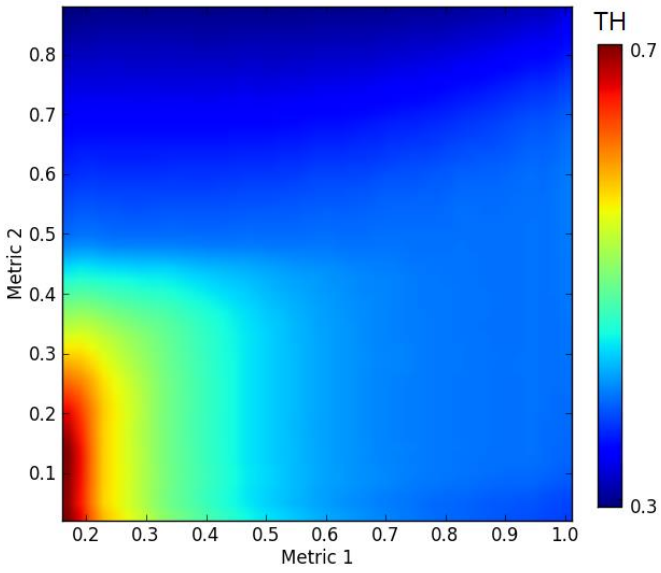


Figure 74. Interpolated result from the metrics coming from Table 11

For using this model over any pattern, first the metrics must be calculated in order to determine which threshold value should be employed. For this reason, this model is significantly slower than the constant threshold approach. Nonetheless, it is faster than the variable threshold presented on previous works since it relies on metrics that are faster to compute (based on kernels). Moreover, its flexibility enables significant precision improvement in some particular cases where more rigid structures fail to explain the e-beam process.

4.4. Conclusion

In the advance of lithography processes, some steps that presented negligible impact in the past are now significantly relevant and must be considered during simulation. One clear example is the acid diffusion in the resist that was usually incorporated as part of the forward scattering model. In more precise approaches it may be included as an independent model, either as a First Principle more rigorous model or as a convolution of a distribution over the aerial image. This is also observed in etch models, which once were considered as a constant bias over the process result but its increasing importance lead to adding more accurate models. The number of different examples is growing as a more advanced technology node is considered.

In this chapter it was discussed the different modeling strategies to predict main effects impacting resolution on e-beam lithography. The advantage of using First Principle Models is that they are based on proved theories from which one may calculate the approximations and errors involved. Moreover, any change in the process may be directly translated into changes in parameter values or even replacing a portion of the model for a different one. However, First Principle Models are as accurate as information about the process is available, which may include sensible and/or proprietary information. Moreover, the use of First Principle models are limited in real situations due to the long time required for each simulation and the complexity of the phenomena being modeled nowadays where more and more different configuration of patterns are required. On the other hand, Compact Models are an alternative

presenting a reduced computational effort for simulation, which is sine qua non condition to be used in production. Moreover, since they are usually implemented over experimental data, they may account for phenomena not fully understood or non-predicted at all. However, using Compact Models requires extra care due to the limitation in the metrology accuracy.

In the context of providing models that may be used on very large volume of data, performance is the main criterion while deciding which model to use. For this reason, a Compact model is usually the tool of choice. Whenever model precision is not a main concern, using a Gaussian PSF model associated with a constant threshold may be sufficient. This is usually the less expensive Compact Model in terms of runtime. However, when it comes to model precision, a good compromise between performance and precision may be employing Pearson PSF model associated with a variable threshold based on the surface model using the advanced metrics.

Nonetheless, the model structure is not the unique factor to impact the quality of a model. Performing a proper model calibration is as critical as working with the most suitable model structure. This is the subject of the next chapter.

Chapter V. Model Calibration

"It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature..."

- *Niehls Bohr*

5. Model Calibration

The model calibration procedure consists of changing values of model input parameters in order to match a reference. This reference may be either field experimental conditions or results coming from a reference model. In this chapter, strategies for calibrating e-beam lithography models are presented.

5.1. Compact Model Calibration for E-beam Lithography

There are two main strategies that one may use to perform Compact Model calibration for e-beam lithography. One is to rely on First Principle Models to provide an accurate description of the e-beam process and then calibrate a Compact Model that attempt to emulate an equivalent behavior. The second strategy relies on performing an experimental exposure followed by measurements and then calibrate a Compact Model that provides results matching the observations. Moreover, it is also possible to combine both strategies, determining some parameters of a Compact Model using First Principle Model simulations and other parameters based on experimental data. Figure 75 shows two examples of empirical model calibration flows, one based on First Principle Model simulations - Figure 75(a) – and another based on Experimental Measurements - Figure 75(b).

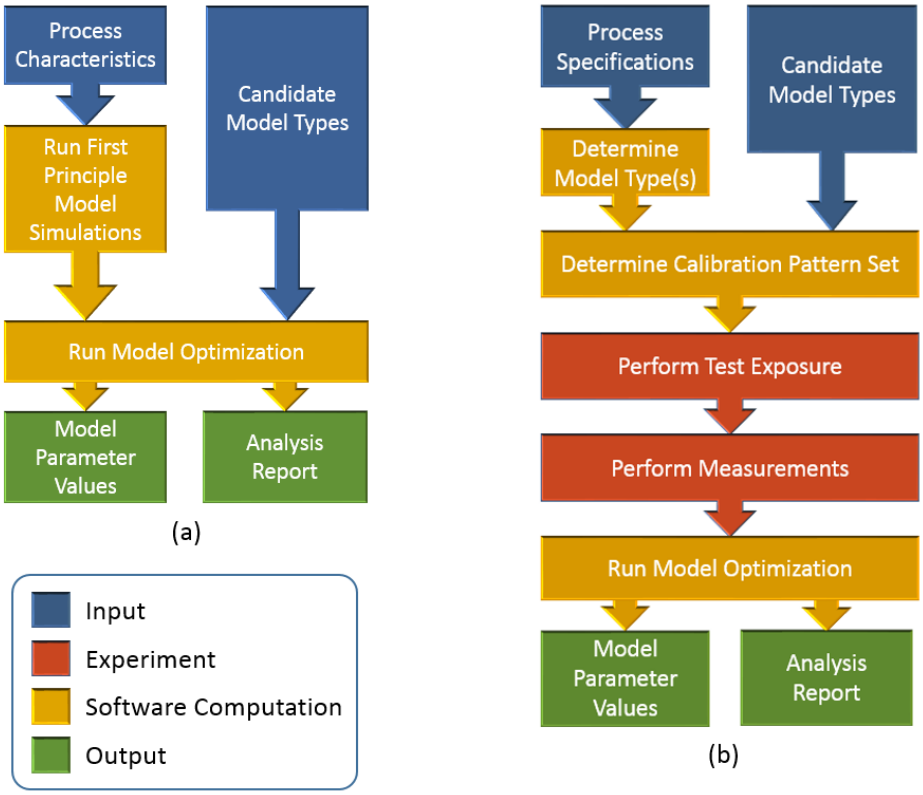


Figure 75. Schematic example of the empirical model calibration flow based on (a) first principle models and (b) experimental measurements.

The first strategy, based on using first principle models simulation as a reference for the Compact Model calibration works as described below.

The process characteristics are studied to be used as inputs to the First Principle Models. This study may be performed either relying on the literature or on experimental results used to calibrate the First Principle Models. These characteristics include all parameters that are not random on the First Principle formulas involved. This strategy is more frequently employed to calibrate the model of a Point Spread Function (PSF) but may be used for evaluating any other step of the lithography process separately.

For the example of a PSF Compact Model calibration, one would rely on a First Principle Model such as the CASINO® software. The Monte Carlo simulation is performed based on the process parameters for a large number of electron trajectories and then the resulting energy distribution in the resist is used as the reference to the model optimization. The optimization algorithm tries to determine the value for each of the Compact Model parameters in such a way that the estimated distribution by such empirical model is as similar as possible to the one coming from the First Principle Model. An example of a 4 Gaussian PSF Compact Model result after calibration based on a CASINO energy distribution is shown in Figure 76.

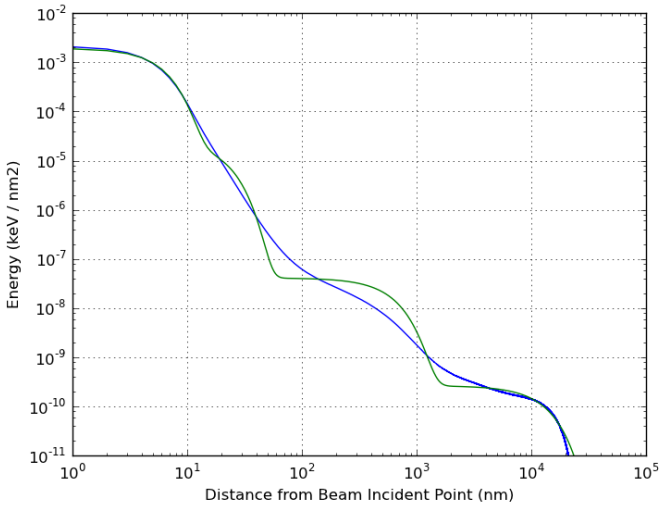


Figure 76. Example of a 4 Gaussian PSF Empirical Model (in green) calibrated over a CASINO energy simulation (in blue).

The second strategy, based on experimental data from exposure and measurements, is usually performed as described below.

According to the process specifications (meaning the targeted minimum CD, minimum pitch dimensions, etc.) one select what model(s) should be considered to be calibrated. Then, a calibration pattern set must be selected and assembled in a design. This is the calibration design, which should be representative of the specifications of the process under modeling and should take into account the model types that are considered for the calibration. Once the calibration design is ready, it must be exposed and measured in exactly the same process flow that is being modeled. Finally, when the measurements

are done one may perform the model optimization, trying to obtain an empirical model that produces the closest values to the measurements possible.

The main advantage of employing a First Principle Model for calibration is that it does not require real experiments if its parameters are known. This is usually a time and money saver in comparison to computer simulation. Moreover, simulations are not affected by metrology variation, providing the Compact Model calibration a noise-free input data. However, this advantage not always works in favor to produce the most precise model. The drawback is that all phenomena that are not integrated on the First Principle Model used are completely disregarded. If experimental data is used instead, the measurements include all phenomena and variability that will impact future exposures. Another advantage of working with experimental data is that the process may be seen as a black-box, not requiring full access to information such as the resist thickness or composition, for instance. For these reasons, the industry usually employs the calibration based on experiments whenever the process is already available in the cleanroom and rely in the First Principle Models only for new nodes which are still under development.

The strategy of performing the empirical model calibration based on experimental data requires not only the exposure and measurement efforts but also the creation of a calibration design. This is the subject of the next section.

5.1.1. Calibration Design

Compact Model calibration based on experimental measurements requires a calibration design to be exposed and measured. The calibration design is usually composed of a series of different patterns that aim to represent the diversity of structures one may find in an actual design.

In order to allow the calibration procedure to define accurately the values of the model parameters, it is important that the patterns are able to stress the characteristics of the process under modeling. For instance, in order to model the range of backscattered electrons, it is necessary to include patterns that present features at different distances, including all of those estimated to be in range and out of range of the backscattered electrons. In this way one will have the information required for calibration.

A simple example is presented to illustrate this concept. Figure 77 shows seven different patterns that may be used to characterize the range (r) of a given effect. The pattern “A” shown in Figure 77 (a) is isolated, which means it undergoes no impact from any proximity effect, but is impacted by electron forward scattering, etching effect, etc. The measurement of Figure 77 (a) may, therefore, be used as a reference for comparing other exposures that are impacted by proximity effects. Figure 77 (b)-(g) show different patterns by changing the distance (d) of the reference pattern “A” to a larger pattern “B”, which is used to induce proximity effects over “A”. As the distance gets smaller, the unknown effect region tends to interfere with the metrology point at “A”, providing a perception of the proximity effect being modeled.

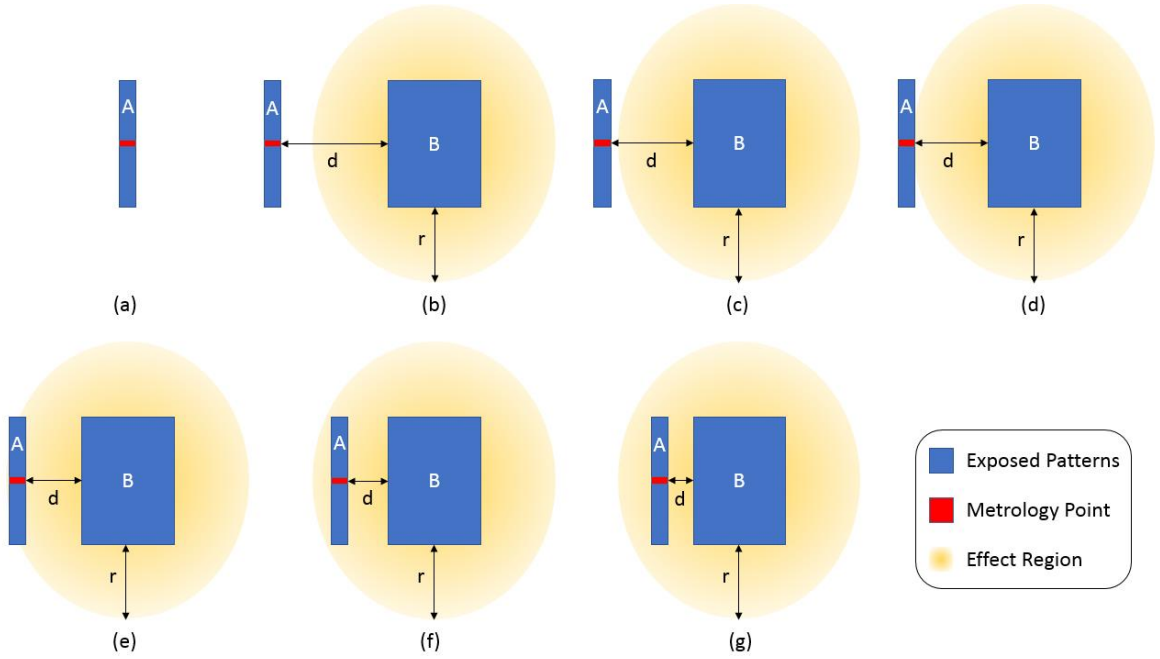


Figure 77. Example of how a pad "B" has impact over the measurement in line "A" according to their distance, varying from (a) to (g).

Consider that these seven patterns are exposed and measured. An example of metrology result is represented as the blue points in Figure 78. A Compact Model may be calibrated in order to generate the closest possible result, as the example shown in Figure 78, the orange curve.

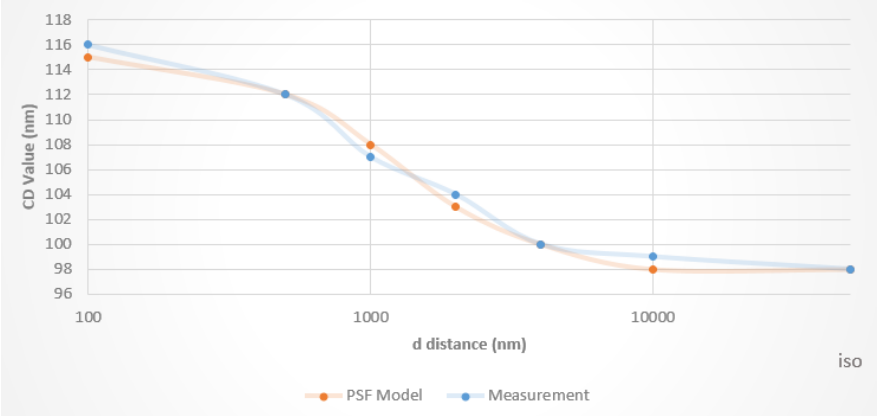


Figure 78. Example of measurement results (blue curve) and Compact PSF Model calibrated to generate equivalent results (orange curve).

In the example presented in Figure 77 it is possible to determine in an intuitive manner that the measurement will be impacted by the variation in distance "d". Approaches based on proposing a simple calibration pattern that presented strong relation to one or several parameters of the PSF were proposed in the literature as much as early as the first correction strategies for e-beam lithography.

One strategy consisted in printing a matrix of circles varying both the exposure dose and the circle radius. The goal is to determine the β and η parameters of a 2 Gaussian PSF that simulate the dose values required to just develop the centers of the circles with different radii, as observed in the experiment [KRATSCHMER, 1981]. Figure 79 shows the calibration layout

and the example of PSF calibrated using this approach. This strategy is not adequate to produce information about the α parameter, though. This limitation was partially overcome by using different metrology tools and relying on the evaluation of the thickness of the material after exposure [RISHTON, 1986]. Nonetheless, this approach is challenging since determining the exact point of development of the center of a circle is difficult.

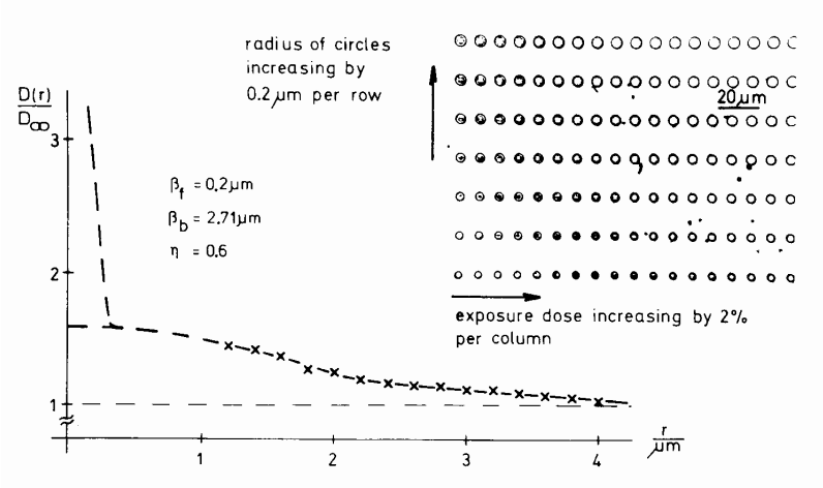
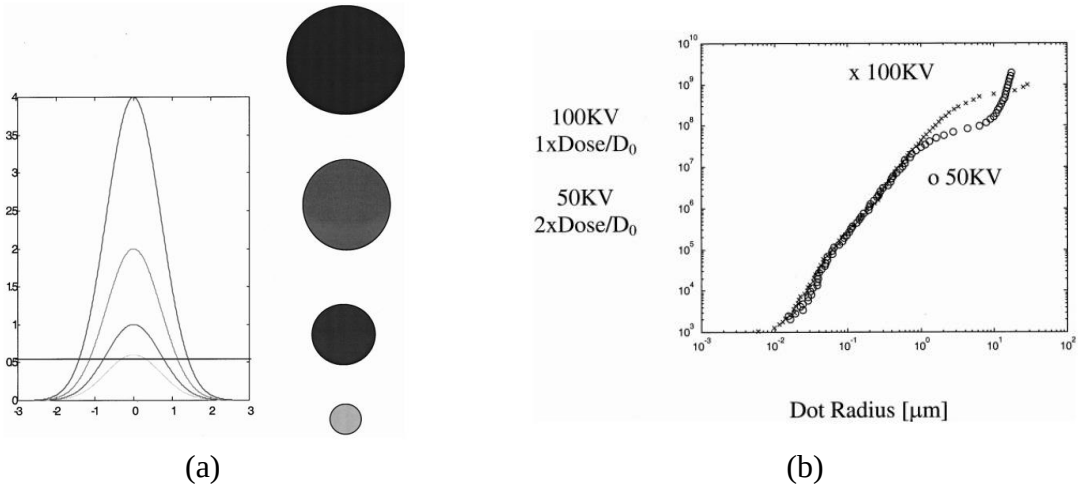


Figure 79. Example of the calibration layout proposed by [KRATSCHMER, 1981]

A variation of this approach using a single circle radius and only varying the dose was also proposed [ANDERSON, 2001]. This approach consists in considering the evolution of the experimental circle width as a function of the exposure dose, as shown in Figure 80(a). This experiment generated a profile as a function of the circle size and the exposure dose (Figure 80(b)).



(a) (b)

Figure 80. (a) Illustration of four different dose profiles and the circles as they would be observed experimentally. (b) The circle (dot) radius as a function of the normalized exposure dose. [ANDERSON, 2001]

A similar approach was proposed some years later which, instead of relying in circles it used doughnut-structures [STEVENS, 1986]. The matrix of doughnuts presented variable inner radius but a constant outer radius.

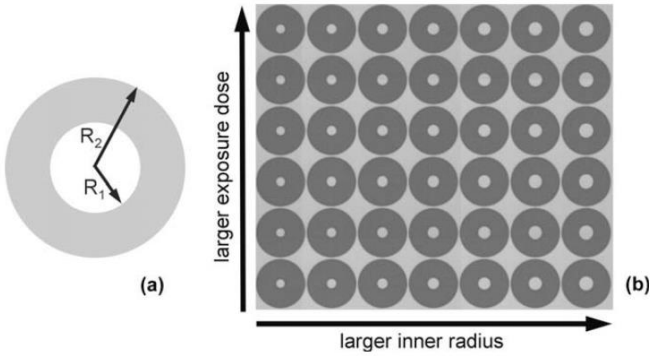


Figure 81. Example of the calibration layout proposed by [HAUPTMANN, 2009]

A more complex pattern was also used by some authors to extend the ability of determining model parameters for Gaussian PSFs with more than two components. One example of such structure is shown in Figure 82 [HUDEK, 2006]. This structure is used to perform a model calibration based on minimizing a cost function and, therefore, may be used to any model for which the parameters are somehow impacted by the proposed structure.

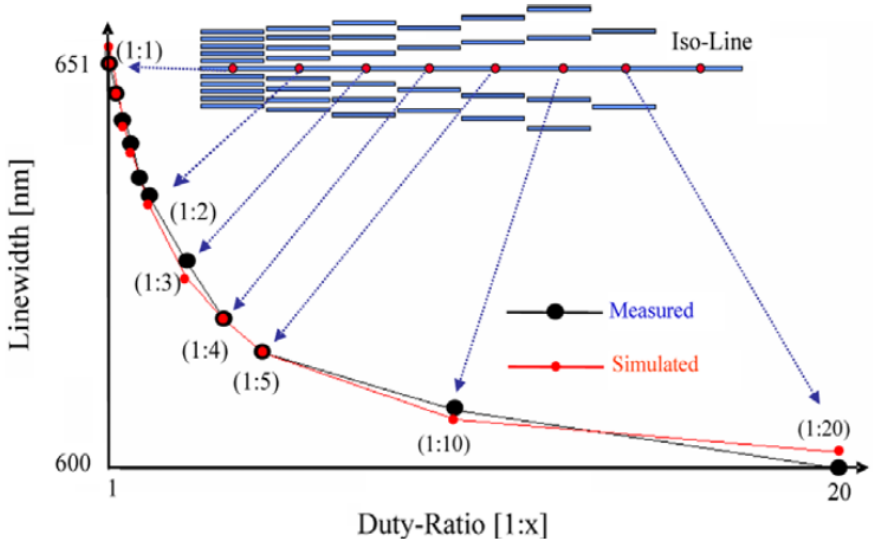
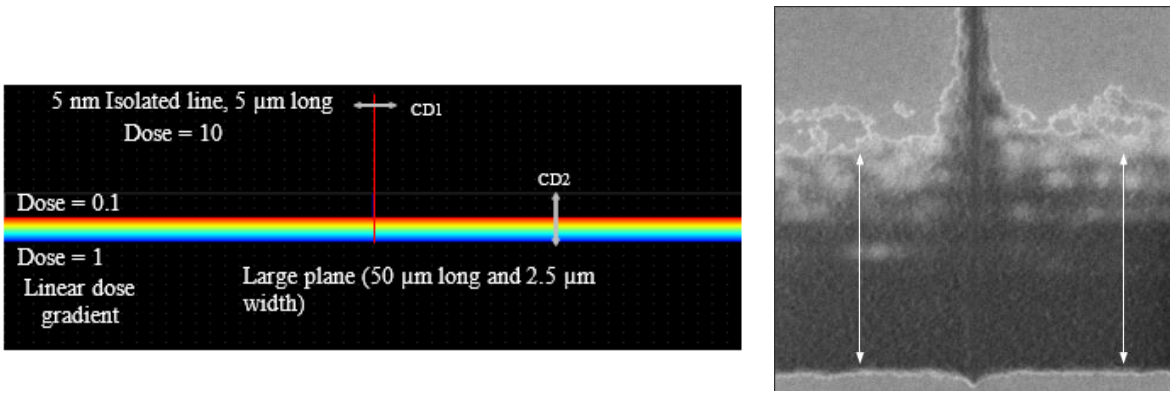


Figure 82. Experimental width of a 600nm line as a function of the density (represented by the duty-ratio of the LSB). [HUDEK, 2006]

Yet another strategy proposed by Babin in 1996 for determining the parameters of a Laser exposure PSF but adapted to be used in e-beam lithography PSF [FIGUEIRO, 2012a]. This strategy is based on using a pattern composed of a very small line exposed using an extremely high dose and a large pattern with a dose gradient. Figure 83(a) shows the test pattern used in [FIGUEIRO, 2012a]. Although one may observe the sensitivity of these parameters to the parameters of a 2 Gaussian PSF, it is complicated to decorrelate to more complex models.

Moreover, the proposed strategy has the constraint of measuring a large plane (which is less precise to measure due to the magnification used in the CD-SEM) and since the dose in one top side of the plane is too low, the variation is significant (this may be inferred from the image in Figure 83(b).



(a) (b)

Figure 83. (a) Example of the calibration layout proposed by [BABIN, 1996] and adapted for e-beam in [FIGUEIRO, 2012a] and (b) the result exposure CD-SEM image.

More recently, a strategy to determine the proximity effect parameters based on CD-linearity has been proposed [HAUPTMANN, 2009]. This approach consists on determining the α and η parameters of a 2 Gaussian PSF by printing isolated lines with widths smaller than the backscattering range. A matrix of patterns varying the isolated line width and exposure dose is used in this experiment. The pair that provides the target dimensions is considered to be the right values for the parameters.

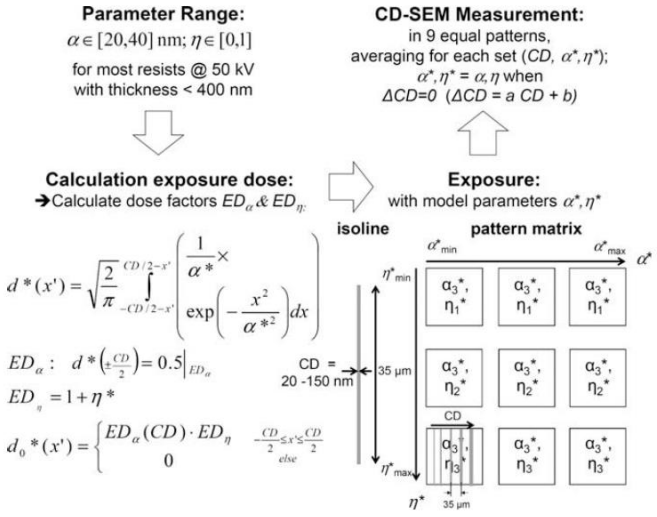


Figure 84. Example of the calibration layout proposed by [HAUPTMANN, 2009]

However, this method is highly sensitive to the process variation. Although it is proposed to perform the measurement over repeated patterns, resist shrinkage, CD-SEM offset and linearity issues may still be an issue [HAUPTMANN, 2009]. In this case a simulation procedure is

proposed by the authors to identify a confidence interval based on the linearity of the CD (which provides a residual standard deviation to the parameter determination procedure).

Another strategy proposed to determine a PSF was presented by [SCHEFZYK, 2010]. In this work an strategy relying in a cross structure. Figure 85 as shown in (. At the center of the cross, the resulting dose is the sum of the contributions of the horizontal and the vertical lines. The overlay of the border lines between fully exposed and underexposed regions of all crosses together helps to determine the values of a function PSF numerically. The main drawback of this approach is the quality of the crosses with low doses and small CDs (smaller than 100nm).

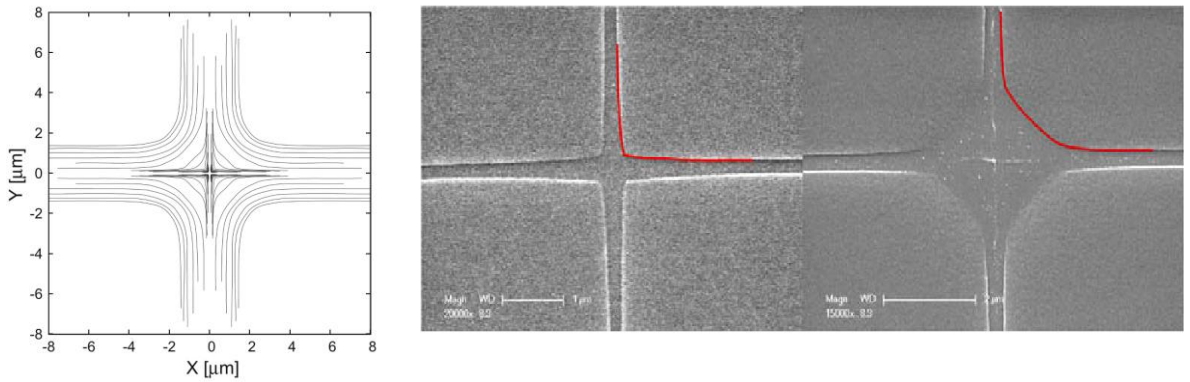


Figure 85. (a) Overlay of the contour lines of several crosses written at different doses. (b) SEM images of two cross patterns written at different doses and the between developed and undeveloped areas. [SCHEFZYK, 2010]

Finally, another approach shown in the literature relies on the "Best Dose Sensor" [UNAL, 2011]. The "Best Dose Sensor" is a specialized calibration pattern that looks like a checkerboard pattern (as shown in Figure 86). The interaction in the corners of the rectangles provide significant information for the short range parameter. In order to improve the sensitivity to other parameters, the pattern was placed in different contexts, as shown in Figure 87.

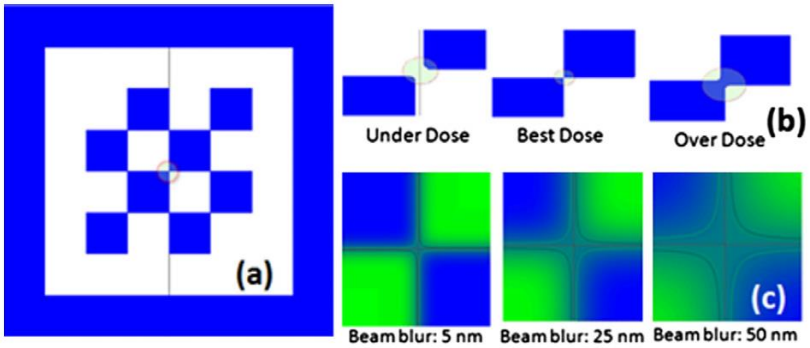


Figure 86. (a) Example of the "Best Dose Sensor" pattern, (b) illustration of its sensitivity to dose changes and (c) the sensitivity of the pattern to ± 10% dose variation depending on short-range blur. [UNAL, 2011].

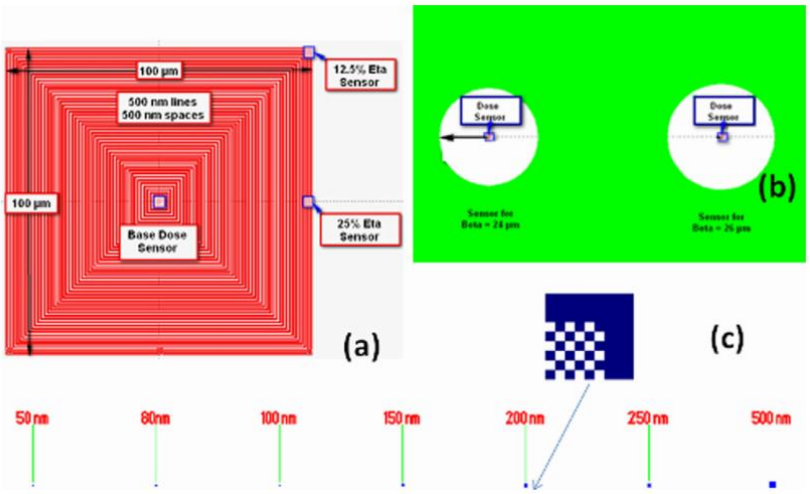


Figure 87. Example of the "Best Dose Sensor" placement for being used to determine (a) base dose and eta value, (b) beta value and (c) mid-range parameters. [UNAL, 2011].

In spite of the large number of strategy available in the literature, an experimental strategy to determine the parameters of a model for e-beam lithography is still a very challenging task. The reason for this is the complexity of the Compact Models used (or required) nowadays for achieving the required precision. Moreover, some of the proposed strategies are limited by metrology resolution or process variations. Therefore, it is useful to have a strategy to verify if a given calibration design present the characteristics required to calibrate with precision a given Compact Model. This is the subject of the next section.

5.1.2. Verifying a Calibration Design

It is possible to state that some calibration designs are more suitable for calibrating a given type of model than others. This is mostly related to the characteristics described previously of presenting patterns that are impacted by the effects being modeled. Considering that an experimental test of exposure and measurements cost time and money, it is interesting to provide a strategy to verify if a calibration design is suitable to the task in hand.

A straightforward way to verify a calibration design is to create an artificial set of measurements of the layout under verification and employ it in several optimization runs. An estimation of the target Compact Model is generated based on the Compact Model structure and parameter ranges that are considered to be used in the calibration procedure. If the calibration is going to be performed using different Compact Models, it may be interesting to perform at least one verification for each model. Once the estimated model is ready, an artificial metrology result is generated based on the simulation of the estimated model over all patterns present on the calibration design. Those results are then used as if they were experimental measurements. Several runs of the model optimization procedure are executed, as if it would be the real calibration. The flow is depicted in Figure 88.

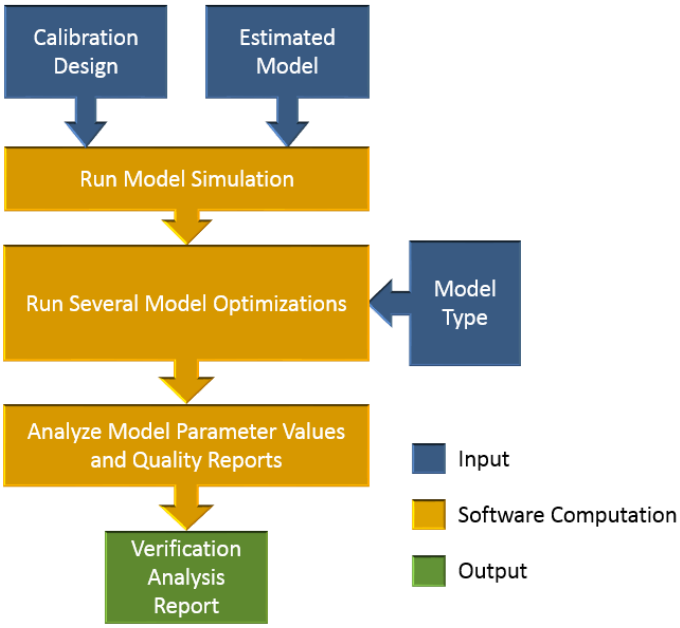


Figure 88. Schematic of a simple flow of calibration design verification.

From the several runs of calibration over the artificial data one may analyze the resulting parameters and cost function values. Ideally, all runs would generate parameter values close to the ones used on the estimated model and, therefore, a cost function value very close to its minimum. This is the scenario where the calibration design passes verification. However, if any of the calibration results provided high errors, it means that the fitting algorithm was unable to find a better solution that existed in the optimization space. Another scenario is that the resulting cost function value is close to its minimum but the parameter values are different from the estimated ones. This indicates that the optimization space presents more than one equivalent solution according to the criteria of the calibration algorithm. This may be prevented by improving the calibration design or changing the cost function. In this case, the calibration design is not verified and another test must be performed either with an improved layout or with a new cost function. Table 12 shows the different results from the several optimization runs and the conclusion one may obtain for each described case.

Table 12. Interpretation of the optimization results of the several runs and its association with the approval of the calibration design.

Result	Conclusion
Some or all error values are high.	The optimization algorithm configuration should be improved and the procedure must be repeated. The calibration design is not verified. ✗
Error values are low, but parameter values vary significantly for the estimated ones.	The calibration design is not adequate to be used for this calibration. ✗
Error values are low and parameter values do not vary significantly.	The calibration design is adequate to be used for this calibration. ✓

For instance, consider a calibration design which is composed only of sixteen isolated lines with a CD varying from 60nm to 200nm. The verification algorithm may be used considering that the target Compact Model is a 2 Gaussian PSF allied with a constant resist threshold. Assuming the parameter values of the estimated model and the range for each parameter as shown in Table 13, one may perform the optimization test.

Table 13. Parameters of a double Gaussian model for verification and the range of variation of the parameters for the calibration exercise also used on the verification algorithm.

Parameter	Reference Value	Minimum	Maximum
α	28	15	40
β	9000	5000	12000
η	0.6	0.1	1.0

The optimization algorithm was performed five times and the resulting parameter values and error (in terms of RMSE) were retrieved. The values are presented in Table 14.

Table 14. Result for five different optimization runs over the Isolated Lines calibration design.

Run	α	β	η	RMSE
1	27.81	7691.1	0.605	0.042
2	27.75	7045.6	0.606	0.061
3	28.05	9273.2	0.599	0.021
4	27.96	9076.8	0.601	0.034
5	27.98	8981.3	0.6	0.02

Notice that the values of the error are low and that optimized values of α and η do not vary significantly, their standard deviation being of 0.13nm for α and of 0.003nm for η. However, the values of β vary over 20%, which is more than what is usually desired. The standard deviation for β is 986.7nm. In this case one may conclude that this is equivalent to the second case described on the Table 12. Therefore, the calibration design is not suitable for calibrating this model. The impact over the β parameter is better illustrated with the profile plot of the energy distribution of the five resulting models, as shown in Figure 89.

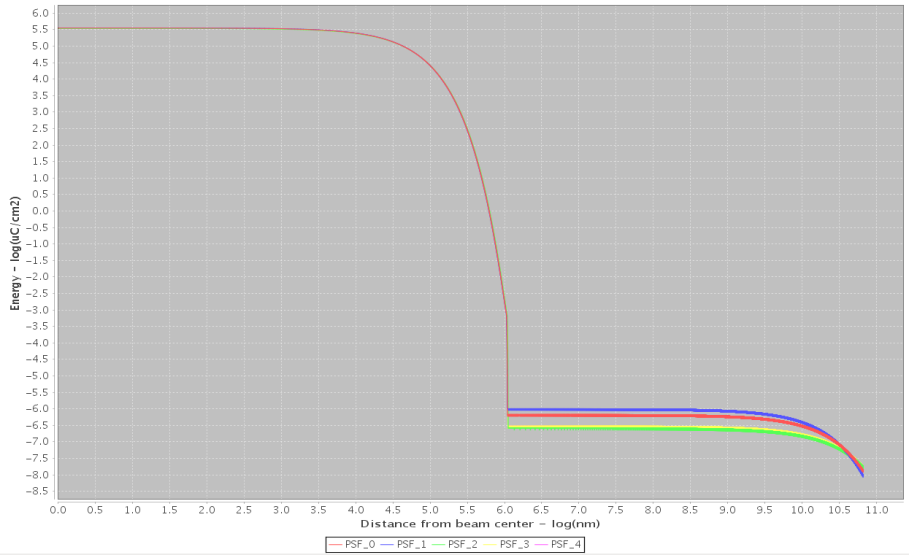


Figure 89. Energy profile plot of the five generated models (in loglog scale) during the Isolated Lines Calibration Design verification.

Observe that on the left hand side of the plot all five models present very similar behavior. This is because this region is represented by the first Gaussian distribution of the model, governed by the α parameter. However, on the right hand size, which accounts for the β parameter, one may see the different behavior for each instance of the 2 Gaussian PSF model.

As shown in the beginning of this chapter, to quantify an effect of proximity between shapes it is necessary to vary the distance between shapes inside and outside the estimated range of this effect. Therefore, a second layout is proposed composed only of line-between-pads (LBP) patterns, which are illustrated in Figure 90.

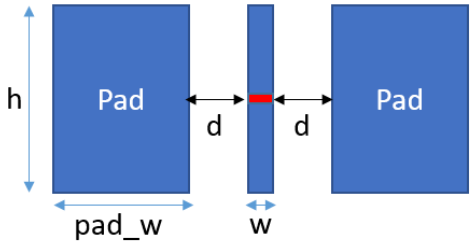


Figure 90. Schematic of the Line Between Pads (LBP) calibration pattern. Where "h" is the height of the pattern, "w" is the width of the measured line, "pad_w" is the width of the pads and "d" is the distance between the pads and the measured line.

A second calibration design is then proposed composed exclusively of LBPs patterns, varying the width of the measured line (w) from 60nm to 200nm (as in the previous calibration design) and varying de distance of the pads (d) from 2000nm to 15000nm.

The same exercise of verifying the calibration layout is performed for this second one, using the same estimated model parameters and ranges (as shown in Table 13). Table 15 presents the resulting parameter values and error (in terms of RMSE) retrieved.

Table 15. Result for five different optimization runs over the LBPs calibration design.

Run	α	β	η	RMSE
1	27.97	8920.3	0.597	0.037
2	28.31	8934.1	0.591	0.039
3	27.74	8920.3	0.601	0.037
4	28.32	8942.5	0.59	0.041
5	27.78	8933.4	0.6	0.04

Notice that the values of the error values are low and that the values of the three parameters do not vary significantly. The standard deviation for α is 0.28nm, while for β it is of 9.65nm and for η it is of 0.005nm. This is a typical case described on the third entry of Table 12. Therefore, the calibration design is verified and suitable for calibrating this model. For comparison with the energy profile of the previous verification (Figure 89), the profile plot of the energy distribution of the five resulting models is shown in Figure 91.

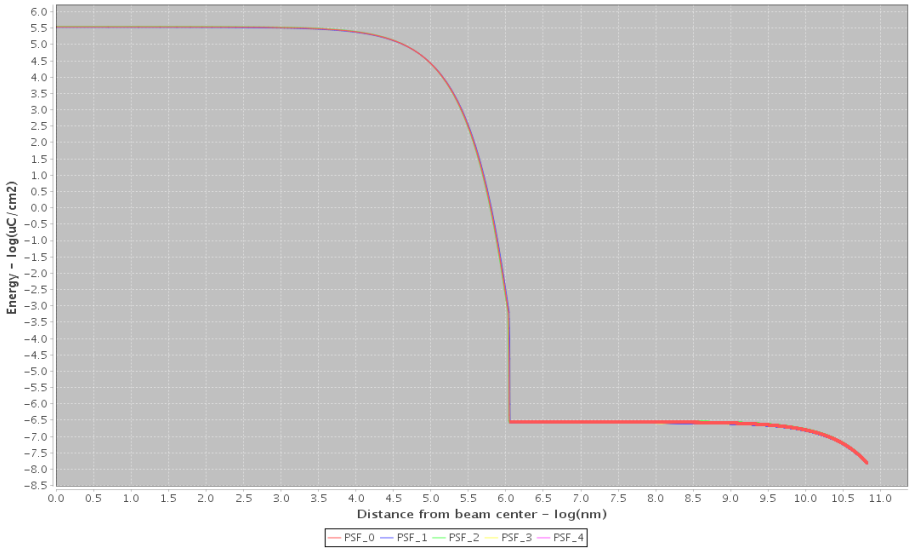


Figure 91. Energy profile plot of the five generated models (in loglog scale) during the LBPs Calibration Design verification.

However, it is not feasible to conceive a calibration design based on trial-and-error approach. In the next section a strategy to build a calibration design is presented.

5.1.3. Building a Calibration Design

Although one may rely on trial-and-error approaches for generating a calibration design, adding different patterns until it passes a verification test, this is usually a time consuming approach that leaves a large room from improvement. Ideally, one should have a solution for building a calibration design that is right by construction.

As described earlier, a calibration design must present patterns that may react to the characteristics of the process that are reflected in the model parameters. This is well represented by the sensitivity analysis indices such as Sobol' 1st Order for non-correlated impact of a given parameter and Sobol' Total Order for both correlated and non-correlated effects. However,

before checking if a given pattern presents the sensitivity to the parameters one is interested in calibrating, it is necessary to known candidate patterns to be evaluated.

5.1.3.1. Candidate Calibration Patterns

In order to determine if any given pattern is a suitable candidate to be evaluated based on sensitivity analysis, one must first evaluate if it is feasible to consider using it in the experimental perspective. For this purpose, some knowledge regarding the process characteristics is relevant in order to prepare a list of realistic candidate patterns.

One of these main aspects is the maximum dimension. CD-SEM tools usually have lower precision when measuring patterns larger than 1µm. Moreover, it is usually necessary to change magnification level from small patterns (up to 200nm) to larger ones (above 200nm). Any change in magnification is an additional phenomenon that impact the metrology results and, therefore, impacts the calibration and must be considered while adding those patterns to the calibration design and during the optimization. Therefore, whenever possible, patterns should be selected in order to be all measured using the same magnification level, unless the target layout requires otherwise.

Another aspect is the pattern density. When exposing very dense patterns, the dose to size (DTS) of such patterns is usually very low compared to the process nominal dose, which leads to a high line-edge-roughness (LER) and low energy latitude (EL). This may increase the uncertainty in the measurement values performed over the calibration design.

Finally, using CD and/or space values that are too small in relation to the process resolution is tempting since those patterns usually are more affected by the model parameters. However, experimentally they are source of uncertainties both in the metrology and in the lithography standpoints. The minimum dimensions to be used should be close to the expected resolution of the process being modeled or close to the specifications of the target layout. For reference, in a process for 14nm technology node, a mask writing process minimum dimension is around 40nm (assist features) and of 60nm (design features). For direct write, dimensions are more aggressive since there is no scaling factor and usually the minimum dimension is close to the technology node reference dimension.

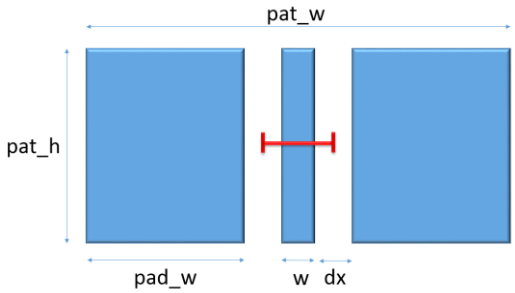
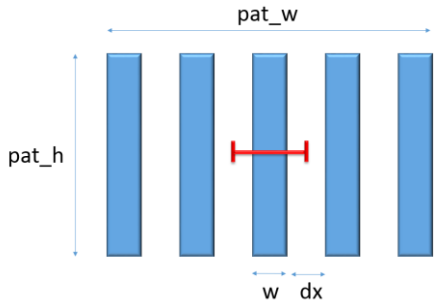
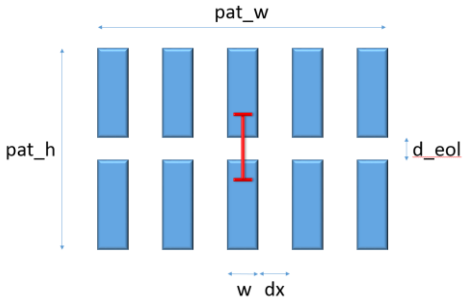
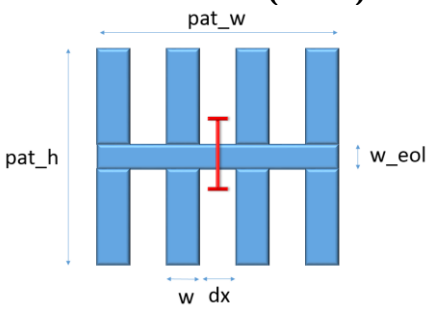
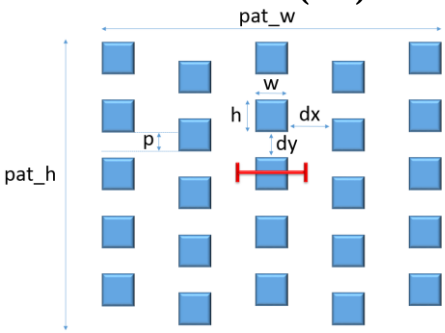
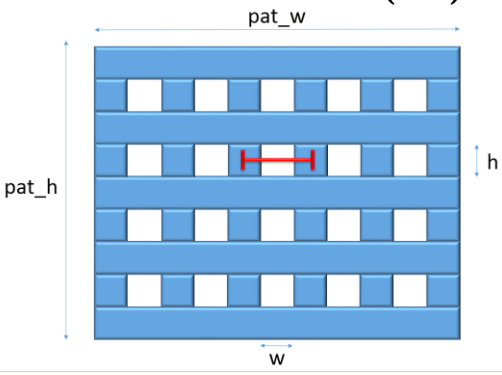
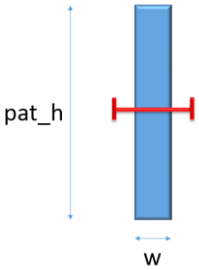
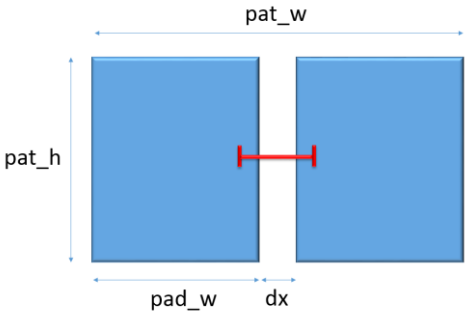
All patterns that passes the experimental feasibility criteria are eligible to be used in a calibration procedure and now it is a matter of selecting the ones that provide more information regarding the process. This is the topic of the next section.

5.1.3.2. Selecting Calibration Patterns

From the beginning of section 5.1.1 it has being stated that a calibration design must be impacted by all parameters of the model. This is a necessary condition to obtain an accurate model calibration. The tool of choice is sensitivity analysis (SA).

The first step of this approach is to have a large database of different pattern types. This database is useful for providing candidate patterns for each of the existing and future models. An example of some pattern types to have in this database and the parameters that may vary are given in Table 16.

Table 16. Examples of calibration patterns types and the dimensions that may vary to improve their sensitivities for different parameters.

Line Between Pads (LBP) 	**L/S Block (LSB)** 
End-of-line Block (EOL) 	**Inverted EOL (iEOL)** 
Dense Contact (DC) 	**Inverted Dense Contacts (iDC)** 
Isolated Line (ISO) 	**Inverted ISO (iISO)** 

The second step of this approach is to determine which of the calibration patterns are more suitable to be employed in a given calibration exercise.

In this work the proposed selection criteria relies on variance based sensitivity analysis. The reason why sensitivity analysis based on the Sobol' indices was selected was already explained in previous chapters but may be summarized by:

- It identifies the patterns whose simulation value are more impacted by the values of the model parameters;
- It allows the evaluation of the contribution of each pattern over each parameter independently (1st Order) and as a combined effect (Total Order);
- It provides a quantified evaluation and comparison among patterns;
- It is independent of any "estimated model";

Another advantage is when it comes to enrich the set of candidate patterns in order to improve results. In a trial-and-error approach, the decision usually relies on the intuitive behavior of the model. This is the case of the example described in the two calibration designs verified in section 5.1.2. However, as the empirical models become more complex, it may be less evident to which characteristics are impacted by each model parameter, making the trial-and-error strategy extremely complex. By relying in sensitivity analysis one may determine for which parameter the calibration design lacks information about and test available patterns to select the one that supply this demand.

In order to compute variance-based sensitivity analysis (SA) over each pattern of the calibration set, the following strategy is performed:

For each pattern candidate to be part of the calibration design, a series of simulations of different models (by varying the input parameters of the model) are performed and the resulting simulated CD is taken as the output of the model. An illustration of this simulation is given in Figure 92.

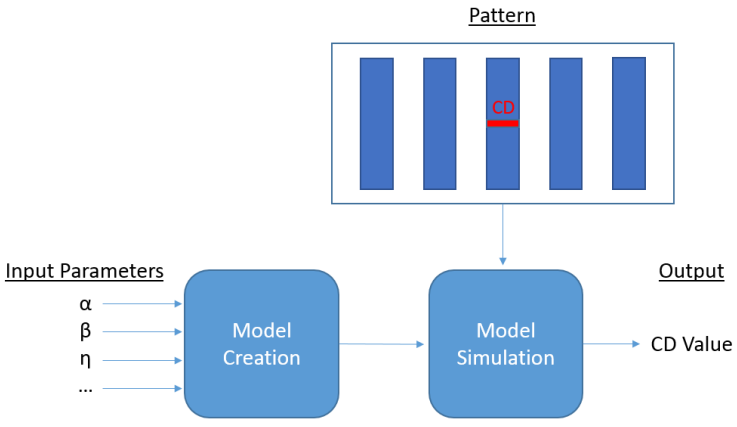


Figure 92. Illustration of the simulation of the CD value of any given pattern

In order to compute Sobol' indices of each input parameter, each input is sampled several times inside the range given for each parameter. The result is how the variation

of the CD value is apportioned to the variation of each of the input values. The estimation of the Sobol' indices is performed as described in section 3.6.5.2.

To illustrate this concept, consider the calibration design composed only of isolated lines, which was used in section 5.1.2. Ranges of variation of parameters are those of Table 13. The Sobol' indices calculated are shown in Figure 93.

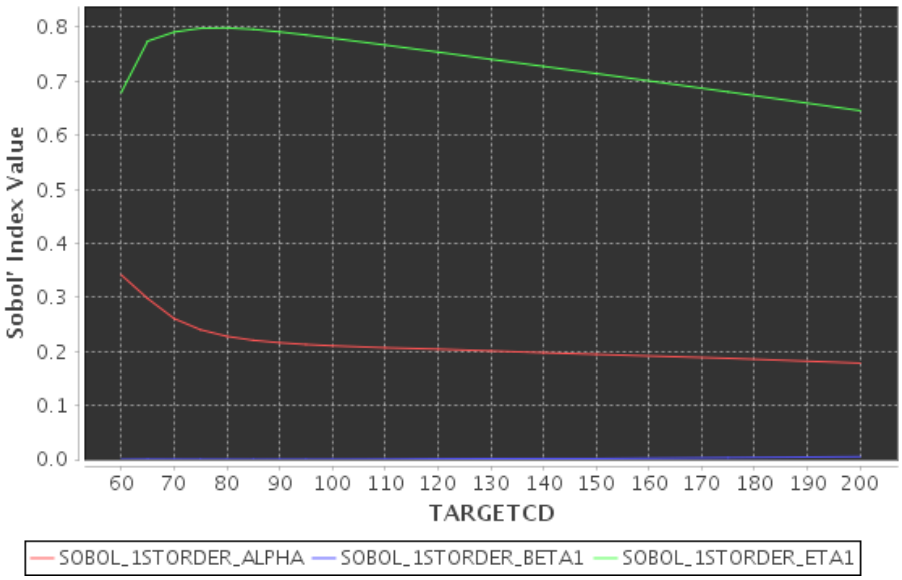


Figure 93. Sobol' First Order Indices for Isolated Lines Patterns

Observe that the layout presents high sensitivity to α and η parameters but poor sensitivity to the β parameter (which may be seen close to zero in the entire interval of target CDs). This is in fully agreement with the result of the verification of the calibration design.

As another example, consider the sensitivity analysis result for the second calibration design used on the verification section, composed of only LBP patterns. The Sobol' indices calculated are shown in Figure 94.

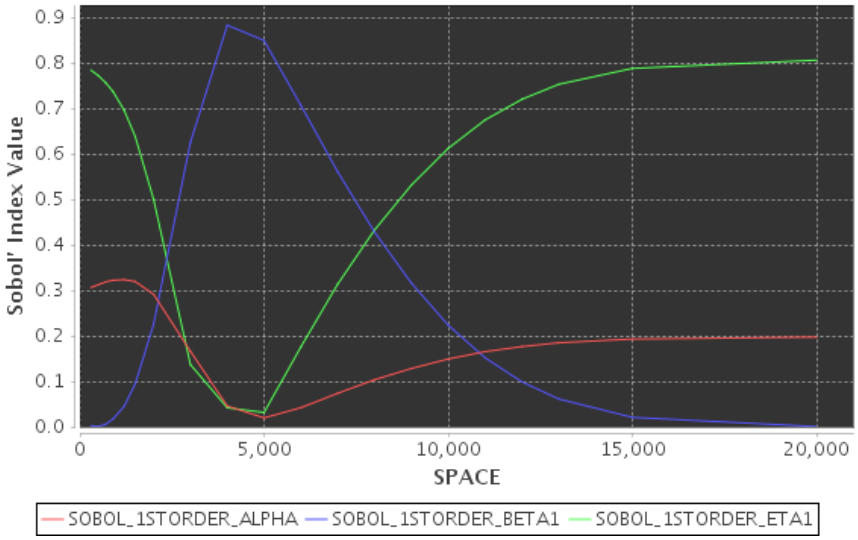


Figure 94. Sobol' First Order Indices for Line Between Pads Patterns with CD of 100nm.

Observe that, according to first order Sobol' indices presented, it is possible to identify that each parameter has a high influence over the simulation result of at least a few patterns. For instance, patterns with small space values present sensitivity to alpha superior of 0.3, meaning that at least 30% of the variance in the simulation may be apportioned to α. Patterns with space around 5000nm present high sensitivity to β and most of the other a very high sensitivity to η. This is in fully agreement with the verification procedure.

Finally, a third calibration layout is proposed. This one includes only three patterns, selected based on the SA calculation. The selected patterns are those presenting the highest first order Sobol' index for each parameter. Therefore, an isolated line of 60nm for α, a LBP of 100nm with space of 4000nm for β and an isolated line of 80nm for η. This would represent the generation of a calibration design with the most suitable patterns a priori. One may verify if it is the case by performing the verification procedure. The resulting parameter values are shown in Table 17 of the five models generated during the verification.

Table 17. Result for five different optimization runs over the calibration design including only the three patterns with higher Sobol' indices.

Run	α	β	η	RMSE
1	27.99	8959.75	0.6	0.0446
2	28.12	8969.33	0.597	0.0426
3	28.28	8972.86	0.595	0.0583
4	28.03	8943.67	0.6	0.0256
5	28.11	8987.6	0.598	0.0775

The resulting parameters generated during the verification exercise over this third layout presented better values than the precedent ones. The standard deviation for α is 0.11nm, while for β it is of 16.28nm and for η it is of 0.002nm. It is important to remark that this result relies on a calibration layout presenting only three patterns.

However, a layout with so few patterns works in a simulation environment because it is not impacted at all by any variability coming from the fabrication process or from the metrology. This subject is discussed in section 5.1.4.

5.1.4. Impact of Process Variation

Whenever models are calibrated based on techniques including one or more cycles of pattern exposure and measurement, it is important to account for variability. First, it is important to quantify the variability. From the e-beam Compact model perspective, it is not relevant to identify the sources of variability since most models do not account for them differently. In this case, the exposure and metrology variability are considered together in the variation on the CD measurement (ΔCD).

One may identify at least three different types of process and/or metrology variability. The first is a distribution (usually Gaussian-like) around a mean value. This is mostly related to small process variations such as the exposure dose, resist thickness, etc. The second one is a bias, either negative or positive, that usually affects all patterns in a layout. This effect may be related to the metrology procedure and the parameters provided to the machine (such as the CD-SEM). The third type is a completely aberrant value that randomly may appear in some of the points. This is actually related to errors in the process, such as a blob in the resist, or to metrology error such as measuring the wrong spot.

5.1.4.1. Variability distribution

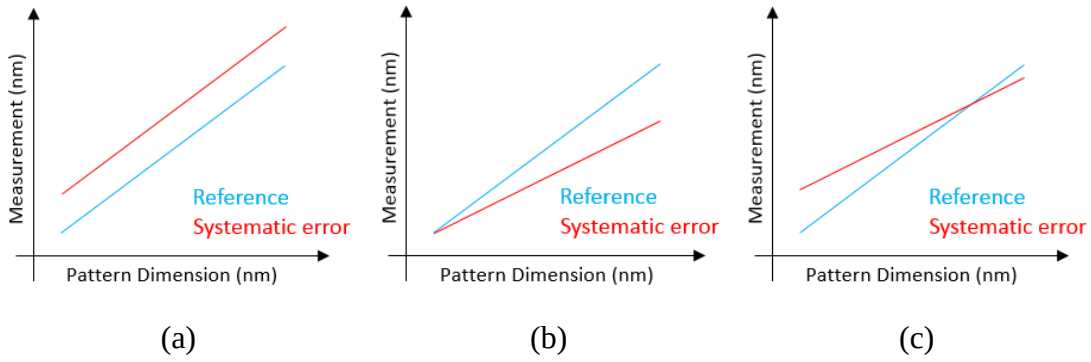
The distribution of points around a mean value is normal and is something found in every process. The concept of CD Uniformity (CDU) and the effort to improve this characteristic is a constant concern in the process fabrication flow because it impacts the minimum features sizes one may use without harming the production yield. This variation is also a major concern for modeling because, since it is related to phenomena that are usually not included on the model (and may be seen as non-deterministic), it provides the precision limits of the model. If the model error is below the variability, it is probable it is fitting the variability (noise). This is a case of overfitting and should be avoided.

5.1.4.2. Systematic Error

Systematic errors over measurements is also common but usually difficult to detect. Metrology systems are usually more concerned in providing precise results than accurate ones and, for this reason, bias is normally detected only when comparing different tools. Other sources of bias may be the use of different e-beam writers, etching tools or any other machine in the fabrication flow, if not properly matched. However, another type of bias one may experiment is when comparing two different exposures on the exactly same flow, usually caused by process variation over time.

One may distinguish two different systematic errors. One is an offset (also known as zero setting error) which are a constant difference between all experimental results. The second one is a scale factor error (or multiplier) which is represented by providing

measurements that are consistently larger or smaller than the reference value by a given factor. An example of how measurements may be affected by systematic errors is shown in Figure 95 for a simple bias represented by an offset factor (Figure 95(a)), for a bias represented by a scale factor as a function of the original pattern dimension (Figure 95(b)) and by a bias which is the combination of the two (Figure 95(c)).



(a) (b) (c)

Figure 95. Example of systematic errors composed of (a) offset factor only, (b) scale factor only and (c) a combination of offset and scale factors.

In either case, systematic errors impact modeling directly. Most of the models will actually absorb this bias which may be dangerous if it is not part of the model structure. It may be classified as an overfitting behavior that may lead to bad predictions when it comes to simulate patterns other than the ones present in the calibration design. This lack of predictability may be solved by including the bias in the model and add patterns to the calibration layout that allow the parameters of the bias structure to be properly optimized.

5.1.4.3. Outlier

The third effect, which may be described as a outlier that replaces the measurement value by a completely aberrant one is not a common event. However, when running hundreds of measurements one may always face a few of those cases. Since they usually happen in small numbers for large measurement sets, calibration algorithms usually are not much impacted by those. Moreover, those points may also be detected and removed from the data using outlier detection algorithms (as described in section 2.1.3). However, if the calibration design is small, every point is important to obtain a good model and a wrong point may mislead the calibration result. Even in the case an outlier detection is successful (which is less likely as the number of points available are small), the missing data may result in a lower quality model.

5.1.4.4. Calibration Design Accounting for Variability

The three different types of process variation described in previous sections must be taken into consideration while building a calibration design. Usually, stable processes have already variability information available. If it is not the case, process variability have to be quantified by experimental exposure and measurement. A layout for such test should include patterns presenting high sensitivity for at least one parameter. This is important because those are the patterns that have influence in the modeling. A

pattern that has low sensitivity to all parameters will not contribute to the determination of the variability of their values and, therefore, is not relevant for this study.

Process and metrology variation are often different according to the type of pattern. One usually observes larger variability on 2D patterns (such as contacts and line-ends) than on 1D patterns (such as L/S gratings and LBPs) [FIGUEIRO, 2013d]. This is usually due to the fact that they are more difficult to print and the metrology has a smaller region for measurement.

Once the different types of patterns are identified and the variations are quantified, one may use a variation of the verification flow described in section 5.1.2 (Figure 88) in order to check if the intended calibration design is robust enough to account for the expected variation on the experiment. This flow is presented in Figure 96.

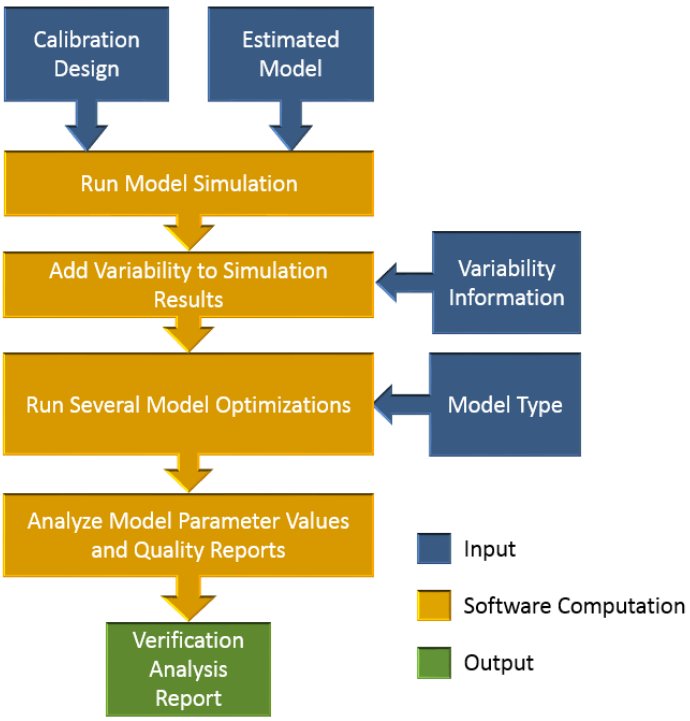


Figure 96. Schematic of a flow for calibration design verification including the variability analysis.

Observe that the difference between the original verification flowchart presented in section 5.1.2 and the one presented in Figure 96 is the additional step of adding variability to the simulated data. The final verification analysis report generated at the end of the flow and the way to read it are identical in both cases.

In order to create the variability on the simulation data, a normal distribution of noise was employed. The choice for this distribution comes from observing distributions of experimental data [FIGUEIRO, 2013d]. The standard deviation of the distribution stands for the process and metrology variability around a mean (section 5.1.4.1). Moreover, a bias value is added to account for systematic errors (section 5.1.4.2). Finally, the original value may be replaced by an outlier according to the probability of its occurrence in the process. The value to be used as a replacer is also randomly

determined. The flowchart for adding the three different sorts of process variation to the set of simulation results coming from the previous step is shown in Figure 97.

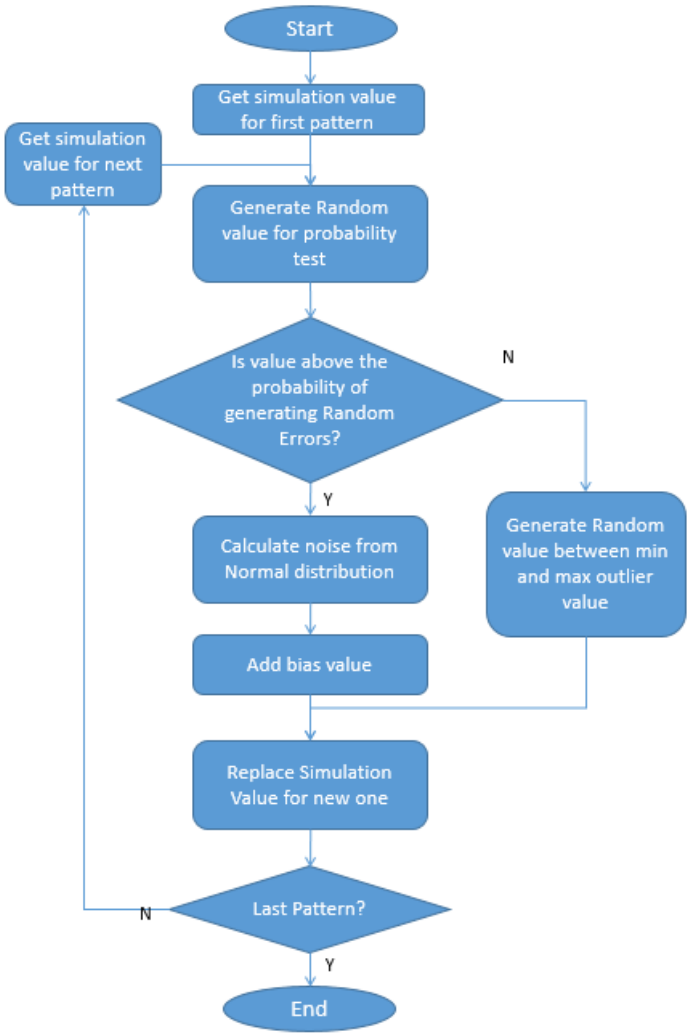


Figure 97. Flow for adding different types of variability to simulation results.

For illustrating the importance of accounting for variability, consider the calibration design containing only three patterns described on the previous section. In this example, only variation is the distribution around a mean value was included, considering a standard deviation of 2nm. The exercise is repeated three times to allow the random values to impact the original simulation in different manners. The final calibration results are shown in Table 18. The ideal values for the three parameters are $\alpha = 28$, $\beta = 9000$ and $\eta = 0.6$.

Table 18. Result for three different optimization runs over the calibration design including only the three patterns with higher Sobol' indices and 2nm sigma variation over the simulation results.

Run	α	β	η	RMSE
1	22,53	9469,94	0,727	0,06826
2	21,15	9894,39	0,757	0,03567
3	28,38	8829,19	0,56	0,0376

Observe that the calibration procedure continues to obtain a solution with low RMSE. This is in fact an issue meaning that the model absorbed the injected variability. Ideally, the RMSE should be equal to the variability used, once that is the error added to the simulation of the target model. The fact that the calibrated model absorbed the added variability is also visible in the large differences between the parameter values depending on the run. This is mainly because the number of test patterns is so low that any single change may induce the optimization to find a completely different model that produces such value.

Observe that if the same test is performed over a calibration layout containing 180 test patterns, being actually the same design with 3 different patterns repeated 60 times, the results are quite different, as shown in Table 19.

Table 19. Result for three different optimization runs over the calibration design including the three patterns with higher Sobol' indices repeated 60 times and 2nm sigma variation over the simulation results.

Run	α	β	η	RMSE
1	28,3	9164,2	0,596	1,9773
2	27,32	9019,2	0,617	1,86123
3	28	8868,1	0,602	2,08915

It is possible to observe that the RMSE values are close to the injected variability (2nm) and that the parameter values are not that far from the target ones. One may continue to add more patterns to the calibration design to continue to enhance the accuracy of the pattern value estimation. On this example the first layout containing only three patterns would be rejected to be employed as a calibration design while this second one could be accepted according to the precision requirements over the model parameters. Therefore, a calibration layout not only must present patterns presenting sensitivity to all parameters of the model but also have them in sufficient numbers in order to account for process variability and random errors.

Another important aspect concerning calibration layouts is to be sure the exposure won't be too far off the exposure targets. This is the subject of the next section.

5.1.5. Calibration Layout Pre-correction

Due to the important of the impact of the effects related to electron-scattering during an e-beam exposure, a calibration layout exposed without any sort of proximity effect compensation will only partially print. For instance, consider the exposure of L/S blocks with very different densities. For the low densities, exposure dose should be higher than for the ones with high densities, otherwise either the first won't have enough dose to print or the second will bridge completely printing as a single pad. Moreover, patterns that print with dimensions very different from their target ones may induce errors in the calibration procedure.

Therefore, some calibration patterns should be exposed using different doses than others. However, it is useful to have similar patterns exposed by the same dose, in order to aid outlier detection. For instance, two LS blocks presenting the same density of 50%, one with a CD of

80nm and another with a CD of 150nm. The patterns are sufficiently different to require different exposure doses to be both in target. However, doing so will provide two different value with two changes (exposure dose and CD value). Those two changes make it more difficult to evaluate the data.

For addressing this issue, we propose applying a simple dose modulation over the calibration design based on an approximate model (such as the one used on the verification exercise). The correction employs a coarse distribution of doses (such as varying from the minimum to the maximum required doses in steps of 10%). By doing so only few different doses are used, reducing the negative impact for the data analysis but enough to print the patterns not too far off their initial target dimension.

Up to this point all strategies regarding the generation of a calibration layout aimed to improve its capabilities for detecting and enabling modeling of effects that impact a single test pattern individually. This is the case for all effects that range from a few nanometers up to several micrometers. However, there are some effects that present ranges way much larger than that, on the order of the millimeter range. For calibrating a model considering these effects a specialized calibration layout is required. This topic is discussed on the next section.

5.1.6. Calibration Layout for Extreme Long-Range Effects

E-beam models are usually employed for correcting "short-range" and "long-range" backscattering effects up to 15µm away from the beam incident spot. It is well known that there are also other process related phenomena that impact pattern uniformity which present a wider range (fogging, chemical mechanical polishing effects, etc.). These longer range phenomena impacts up to a few millimeters [KAHNG, 1999][HUDEK, 2007]. These effects may be called "extreme-long-range" for distinction from the 10µm to 15µm "long-range" effects.

The calibration layout should be able to evaluate the importance of the density and its impact over patterns that are placed over a large distance. The strategy adopted is based on adding a region that generates the extreme long-range effect and some patterns to probe the phenomenon and suffer its influence [SUNDBERG, 2010], [FIGUEIRO, 2012b]. An example of such layout is presented in Figure 98.

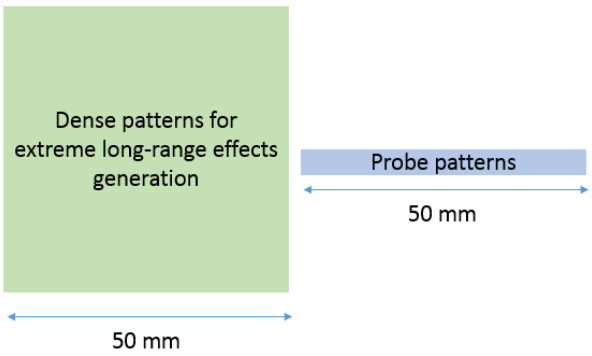
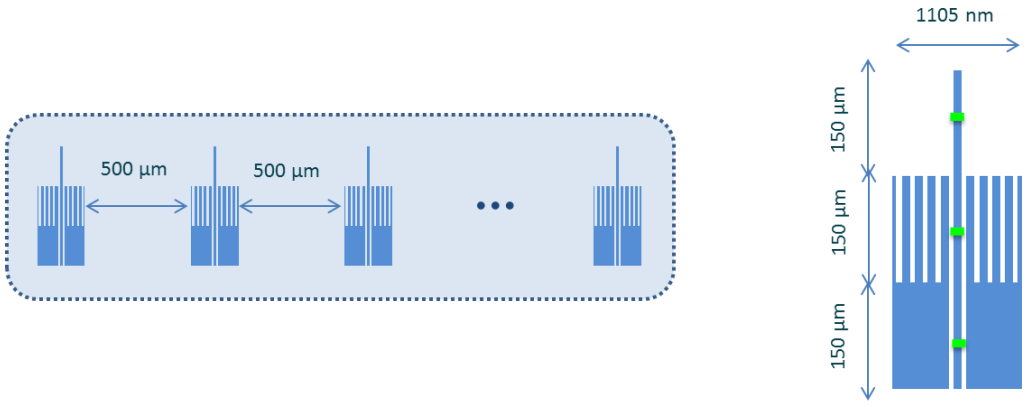


Figure 98. Example of a generic extreme long-range calibration design presenting the two separated regions, one for generating the extreme long-range effects and the other to detect and quantify such effects.

The portion responsible for detecting the effects is very similar to any other calibration design. Ideally it should contain a pattern that suffers little impact from process and metrology variability and should present different densities so both effects may be decorrelated. An example of a set of probe patterns is shown in Figure 99 (a) and one of them in more detail in Figure 99 (b).

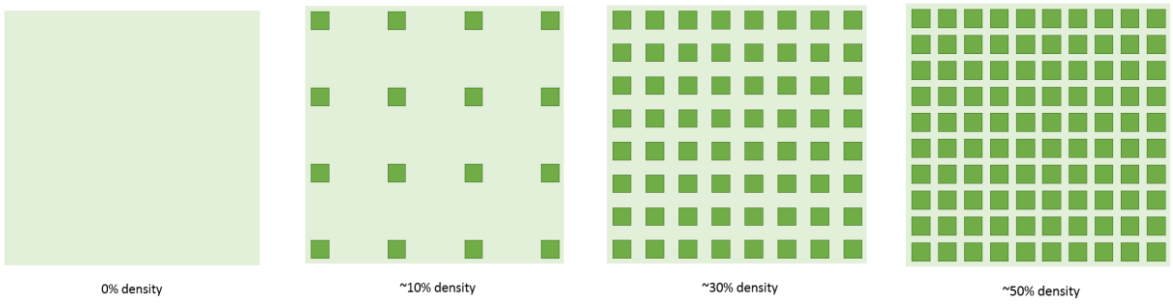


(a) (b)

Figure 99. (a) Example of the set of patterns to be exposed and (b), detail of one of the test patterns [FIGUEIRO, 2012b].

Observe that the test pattern used is composed on only three features to measure, which is either an isolated line (ISO), the center line of a LSB and a LBP. The width of the line should be above the critical dimension of the process under evaluation, so a small variability is observed. Moreover, adding three different densities helps decorrelating local density effects from the extreme-long-range ones.

The portion responsible for generating the long range effect is usually a large block of patterns (either lines or contacts). In order to calibrate the effect, it is the density of this large region that must vary. An illustration of these blocks is given in Figure 100.



(a) (b) (c) (d)

Figure 100. (a) An illustration of the portion of the extreme long-range effects responsible for generating such effect presenting densities of (a) 0%, (b) 10%, (c) 30% and (d) 50%.

Due to the large area required to sample a single density, it is often the case where each density block and the patterns to detect its effects are exposed in different wafers/masks. If that is the case, the inter-exposure variability impact must be considered in order to obtain a more accurate model. Usually this may be performed by adding, in a distance region of the same

wafer/mask, a process variability check pattern containing just a few patterns if high sensitivity repeated several times in order to compare different exposures.

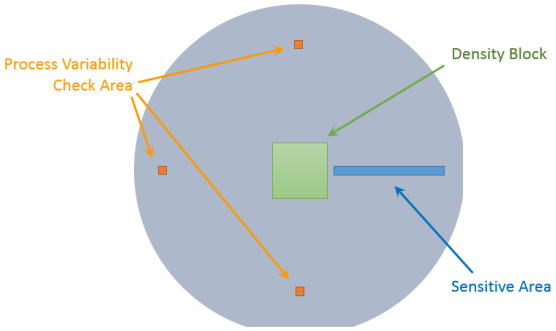


Figure 101. A representation of a wafer floorplan ready to test the impact of a given density over the calibration pattern. Process variability is evaluated by comparing the results combing from the measurements of process variability check area from different wafers.

5.2. E-beam Lithography Model Validation

Once the model is calibrated, it is important to validate it in order to assure that it is an accurate enough representation of the lithography process. This can be performed in different ways.

One strategy is to rely in a new experiment. A correction is performed over a validation design using the generated model. The corrected layout is exposed and measured. The main advantage of this method is that it is very close to the real use of the model and also includes the impact of using different exposures. The drawback of this approach is that it requires a correction to be performed, followed by the exposure and metrology procedures, which results are costly and time consuming.

A second strategy is to include extra patterns on the calibration layout, or to not use some of the patterns available in the optimization process. The extra patterns (or the unused ones) are kept to be employed exclusively in the validation process. In comparison to the previous approach, this one does not require neither correction to be executed nor an extra exposure. As a negative aspect it does require extra metrology effort which do not add quality to the model.

Finally, a third approach is to use the same patterns for calibration and validation, in a cross-validation approach. This approach has as advantage the fact that no extra exposure is required and all patterns contribute to the model quality. On the other hand, it requires extra effort for the calibration process, since it is performed several times.

Usually, it is more interesting to combine these strategies. Since verifying the model through another exposure and measurement after correction present a high cost both in terms of time and money, it is reasonable to employ a first validation based on cross-validation strategies. This strategy relies uniquely in data already available and it is usually a strong evidence of the model quality. If the model passes this first validation test, then it may be used for correction of a different design. The design to be used in this experimental validation should include both patterns that were present in the calibration design and others that were not. The first group is

used to evaluate the model ability to describe the phenomena that were present during the calibration while the second is useful for determining its predictability over new patterns.

If a model fails to pass the validation test, it may represent that the model used does not present the adequate structure to explain the phenomena under modeling. Also, it may be that there is not enough data to perform an accurate calibration due to process variation (maybe the estimation used for verifying the calibration design was not accurate enough). Finally, it may be the case where the data contains too many outliers and requires an extra effort in improving it.

5.3. E-beam Lithography Model Selection

As previously described in section 2.5, there are several different strategies to select the most suitable model among several candidates. For the purposes of modeling electron-beam lithography we found more suitable to rely on a cost function, such as MSE or MSLE, as a first criterion of model quality associated with a coefficient of determination, which indicates how the variance of the data is explained by the model. Specificities of these two approaches for the e-beam lithography modeling are described in the following sections.

5.3.1. Cost Function for e-beam modeling

The selection of the cost function for calibration of an e-beam lithography model has a significant impact on the resulting model parameters. For instance, if one uses MAE cost function, there is no extra weight for large errors in comparison to small ones, making ten errors of one nanometer equivalent to one error of ten nanometer. If one uses MSE instead, which accounts errors in a quadratic way, it would take one hundred of one nanometer errors to compensate a single one error of ten nanometers. For the purpose of model calibration for e-beam lithography, where small errors in simulation due to process variation are common even on the best of the models but large errors should not be common (especially if an outlier detection algorithm was used), MSE seems to be the most adequate cost function.

When it comes to model comparison, one of the first metrics is to compare their final value based on the cost function metric after calibration. This is an interesting approach to compare models when the goal of such model is to explain the data it is calibrated over and predict future events in similar conditions, such as it is the case for e-beam model calibration based on experiments and not so much for the e-beam models calibrated based on Monte Carlo electron scattering simulation.

Moreover, one may also find relevant to observe the results that generate the minimum and maximum errors, both restricted and not restricted to the ±1.5IQR criteria, in order to obtain a clearer view of the behavior of the models being selected. Nonetheless, these criteria do not provide a direct comparison between models of different complexity. For this purpose, we considered the use of the Adjusted R^2 the more pertinent approach.

5.3.2. Coefficient of Determination for e-beam modeling

As stated in section 2.5.2.3, the coefficient of determination (R^2) provides the information regarding how much of the experimental data variance is explained by the model. Its advanced variation, Advanced R^2, also takes into account the number of parameters in the model being compared. However, the approach as originally described [THEIL, 1961] is not directly suitable for our purposes because it relies on the amount of explained variability in relation to the complete variability of the experimental data (in our case, mainly measurements). Since the original intended dimension of the calibration pattern itself is the most significant factor to determine the measurement result, R^2 is biased by it. Therefore, it is interesting to adapt this approach to take the target dimensions into consideration. This is the proposition of the R^2 from Target, or R^2T.

The proposed approach is to no longer evaluate the amount of variability of the experimental data is explained by the model but rather the amount of variability of the deviation from target present in the experimental data is explained. The computation is performed as shown in (90).

$$R^2T = 1 - \frac{\sigma^2_{resToTarget}}{\sigma^2_{Target}}, \tag{90}$$

Where $\sigma^2_{resToTarget}$ is the variance of the residual of the model in respect to the target (and that equals σ_{res}^2, as shown in (91)) and σ^2_{Target} is the variance of the data according to its respective target (as computed in (92)).

$$\sigma^2_{resToTarget} =$$

$$\frac{1}{N}\sum_{i=1}^{N}((m_i - tgt_i) - (x_i - tgt_i) - \mu((m - tgt) - (x - tgt)))^2 = \tag{91}$$

$$\frac{1}{N}\sum_{i=1}^{N}((m_i - x_i) - \mu(m - x))^2 = \sigma^2_{res}$$

Where tgt_i is the target CD value for the i^{th} point, m_i is the measurement value for the i^{th} point, x_i is the simulated value for the i^{th} point, and $\mu((m - tgt) - (x - tgt))$ is the mean value of the difference of the mean of the measurements minus targets and the simulation minus targets.

$$\sigma^2_{Target} = \frac{1}{N}\sum_{i=1}^{N}((x_i - tgt_i) - \mu(x - tgt))^2 \tag{92}$$

Where tgt_i is the target CD value for the i^{th} point and $\mu(x - tgt)$ is the mean of the differences between target and measurement. Therefore, the equation presented in (90) may be calculated as shown in (93).

$$R^2T = 1 - \frac{\sum_{i=1}^{N}(m_i - x_i - \mu(m-x))^2}{\sum_{i=1}^{N}((x_i - tgt_i) - \mu(x-tgt))^2} \tag{93}$$

Notice that essentialy R^2T is not a new metric, but a different interpretation of R^2. Instead of considering the variance as a whole, R^2T evaluates the amount of variance around a target value that is explained by the model. One may also consider the Advanced R^2T to be implemented

by replacing R^2 by R^2T in the original Advancer R^2 equation. The resulting equation is shown in (94).

$$Advanced\ R^2T = 1 - \frac{N-1}{N-(k+1)}\ (1 - R^2T) \tag{94}$$

In order to illustrate the advantage of using such technique consider the following example. Consider that a set of 12 different calibration patterns are exposed and measured. The measurement results are used to calibrate three different models, Model 1 presenting 10 parameters, Model 2 presenting 2 parameters and Model 3, presenting 0 parameters. All values of this example are shown in Table 20.

Observe that the Model 1 presents the best results with an RMSE of 2.3094 nm. The second model presents a RMSE of 7.5166 nm and is worst or equal to Model 1 in every single point. Finally, Model 3 is a copy of the target CD, which would represent a model that predicts that everything would print perfectly at the target. The RMSE for Model 3 is 29.2389 nm.

Table 20. Target CD, measurement and simulation result for three different models for 12 different calibration patterns.

Target CD	Measurement	Model 1	Model 2	Model 3
40	35	35	32	40
40	41	41	45	40
40	42	44	47	40
60	58	59	62	60
60	62	63	66	60
60	71	72	75	60
800	750	752	755	800
800	810	813	819	800
800	880	885	898	800
200	190	191	200	200
200	210	213	215	200
200	230	233	235	200

Using the formulas described in section three, we obtain the following values for R^2 and R^2T as shown in Table 21.

Table 21. R^2 and R^2T values computed for the three models.

	Model 1	Model 2	Model 3
R^2	0.99998	0.99977	0.99183
R^2T	0.9623	0.82409	0

Observe that R^2 considers that all three models explain at least 99% of the variance of the experiments, which is explained due to the fact that all models include in some level the variation of the Target CD. This is well illustrated by the result of Model 3, which only accounts for the variability of the Target CD. On the other hand, the metric R^2T is able to eliminate the influence of the Target CD and identify only the significant variance for evaluation. R^2T results

show that Model 1 explains around 96% of the total variance excluding the different targets, while Model 2 explains 82% and Model 3 explains 0%.

Although no longer presenting its original meaning, R^2 is able to identify that Model 1 is better than Model 2 and that both are superior to Model 3. Nonetheless, this difference has significant impact when one consider the Advanced R^2 and Advanced R^2T indices, as shown in Table 22.

Table 22. Advanced R^2 and Advanced R^2T values computed for the three models.

	Model 1	**Model 2**	**Model 3**
Advanced R^2	0.99978	0.99969	0.99183
Advanced R^2T	0.58535	0.75812	0

Notice that Model 1 has 10 parameters while Model 2 has only 2 parameters. Nonetheless, the Advanced R^2 continues to indicate Model 1 as preferable. Contrarily, the Advanced R^2T index identify Model 2 as significantly better than Model 1.

Therefore, we use the Advanced R^2T criteria in order to compare different models with different number of parameters, allied with the cost function. The Advanced R^2T criterion is not sufficient by itself because it is concerned about explaining the variance, but not the average error. A second simple example illustrates the fact. Consider a Model 4 and a Model 5, which are both calibrated using the same measurements presented in Table 20. Consider that Model 4 performs a perfect calibration, with every simulation identical to the measurement while Model 5 has a positive bias of 10nm for each measurement. The simulation results for this two models are shown in Table 23.

Table 23. Target CD, measurement and simulation result for three different models for 12 different calibration patterns.

Target CD	**Measurement**	**Model 4**	**Model 5**
40	35	35	45
40	41	41	51
40	42	42	52
60	58	58	68
60	62	62	72
60	71	71	81
800	750	750	760
800	810	810	820
800	880	880	890
200	190	190	200
200	210	210	220
200	230	230	240

The values of R^2, R^2T and their advanced variants are all identical and equal to one for these two models. The result is not a surprise since the variance is entirely explained by both models and the mean error is not taken into consideration. Therefore, two conclusions may be derived from this result. The first is that Model 4 (perfect fit with the measurements) and Model 5 (a biased model by 10nm, giving a MSE of $100nm^2$) are equally good by any criteria based on the

coefficient of determination. The second is that Model 1, which had an error of 5.33nm², is considered inferior to Model 5 and its MSE of 100nm².

5.4. Conclusion

Model calibration for e-beam lithography is a critical step in the entire e-beam lithography flow because it is the quality of the model used for PEC one of the major aspects to limit process resolution. In order to calibrate a Compact Model based on experimental data it is key to plan the experiment properly. Relying in a calibration design presenting high sensitivity indices for all parameters is a shortcut to obtain accurate models with reduced metrology effort. Moreover, accounting for process and metrology variability allows planning the number of patterns in the calibration set according to the desired precision of the model parameter values.

In what concerns model validation, it is the last stand to ensure that the generated model is reliable. By employing techniques such as the cross-validation approach, one may validate the model without requiring an extra set of data. The strategy prevents dedicated exposure and measurement only for validation purposes, allowing the use of more data for calibration and, at the same time, as it includes the calibration set as part of the validation set, increasing the data available for validation as well.

E-beam model selection after validation is the last step of the modeling flow. Selecting the most suitable model for the task requires a compromise between model accuracy and complexity so large tasks such as full chip PEC are performed in a reasonable time. In our experiments it is usually selected the model that presents a MSE inferior to a certain threshold (usually related to the process variability or specification) that presents the best Advanced R^2T index.

Chapter VI. Other Applications

“Once we accept our limits, we go beyond them.”

- *Albert Einstein*

6. Other Applications

In the course of this thesis work some developed techniques and understanding of the industry needs lead to the development of other applications. In this chapter one of the most representative is presented.

6.1. Extreme-long-range Effects Correction

Extreme-long-range effects are a significant factor of uniformity errors in current technology nodes and its importance is increasing from one generation to the next. Whether it is CMP, etch loading, resist loading, or e-beam fogging, this effect impacts the entire surface of the reticle or wafer. We propose in this section a technique to compensate extreme-long-range effects going across chip boundaries to the full reticle exposure field (for the case of mask writing) or for the full wafer (for the case of direct write).

As presented in section 4.3.2, extreme-long-range effects may be predicted by using a Compact model such as a simple Gaussian distribution. However, due to the range of several millimeters up to a few centimeters of these effects, performing proximity effects corrections can result in lengthy run times as file size and pattern densities continue to increase exponentially per technology node. Running corrections for extreme-long-range phenomena becomes computational and file size prohibitive. Nevertheless, as process requirements become tighter and tighter, full field density analysis accounting for scribe frames and all neighboring patterns is required for reaching fidelity control requirements such as critical dimension (CD) and line end shortening (LES) on the full plate. Hence, the problem is to perform the complete analysis minimizing the computational effort [FIGUEIRO, 2013a].

The way to address the computation of such large regions without a large computation effort is to rely on pattern density. The entire reticle field's density on the case of a mask write or the entire wafer density on the case of direct write may be computed using surface area density percentages and pixelated per the user's granularity. Once an entire surface density is established, a process specific PSF model may be evaluated over the resulting density and the resulting impact may be taken into account along with traditional short range and long range PEC methods for full field correction.

In order to compensate the impact of extreme long range effects over the entire reticle or wafer, the following steps are proposed and illustrated in Figure 102. Notice that the procedure consist of mainly two modules, one implementing the model calibration procedure (Calibration Module) and the other implementing the correction flow (Correction Module).

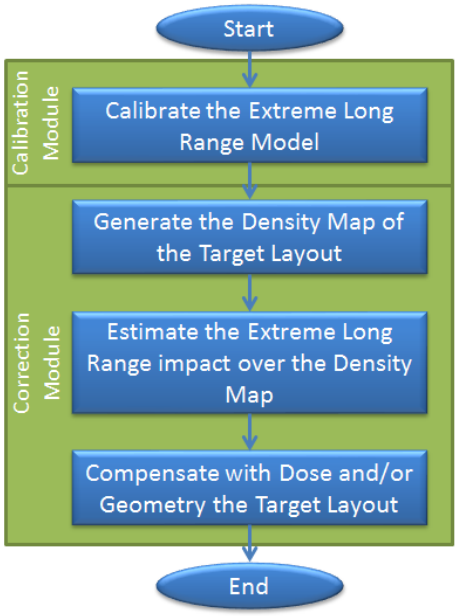
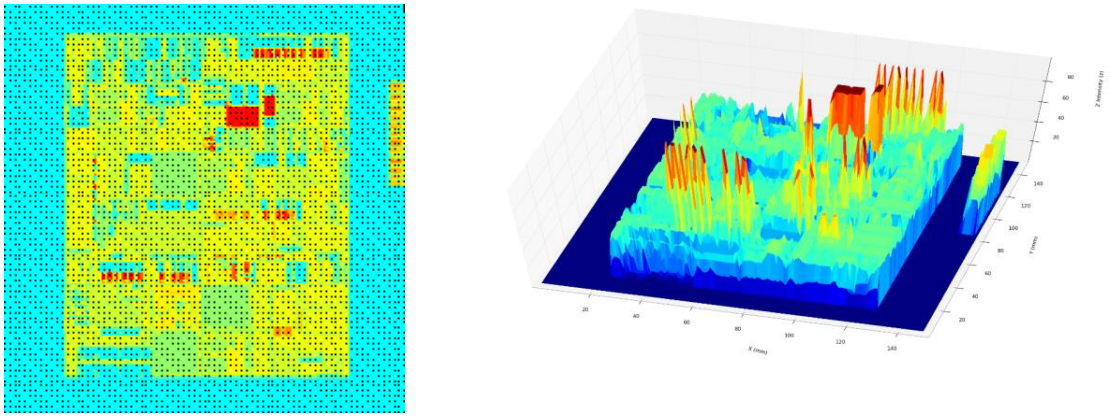


Figure 102. Proposed flow for extreme long range effects compensation.

Step I - Calibrate the Extreme Long Range Model: This step should be performed as described in 5.1.6.

Step II – Generate the Density Map of the Target Layout: The target layout must be evaluated in terms of density. The proposed approach consist in no longer consider each shape individually but the impact of several shapes at once instead. This is accomplished by the extraction of the surface density for the entire reticle. By partitioning the surface regions into manageable pieces and re-assembling the density percentages into a density map, the task can be handled rather efficiently. So calculating the density on a customized tile size area and storing that into a map becomes much quicker and easier to process. Figure 103 shows a density map representing an entire reticle field that will be later applied with a PSF to the individual pattern file. The density pattern shown in Figure 103 contains is divided in pixels of 100µm x 100 µm. By using pixels to evaluate the density it is possible to evaluate an entire wafer of 300mm using around 7M pixels.

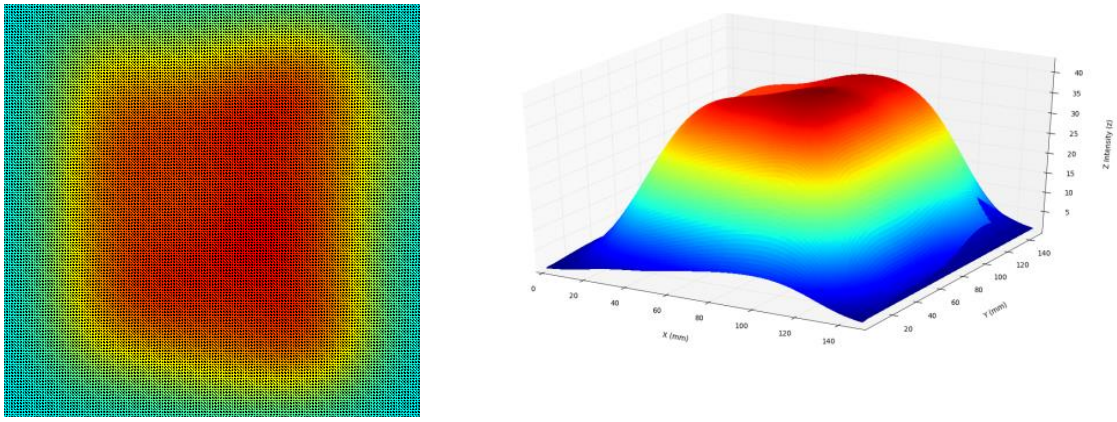


(a) (b)

Figure 103. Density pattern over entire 150x150 mm2 reticle field in (a) 2D plot, (b) in a 3D plot

Step III – Estimate the Extreme Long Range impact over the Density Map: Once the model is calibrated and the density map is available, the convolution of them both will indicate the predicted impact in dose distribution related to extreme long range effects over the entire reticle. The result of this step is a weighted density map [ABE, 2007], which indicates the background dose distribution induced by the extreme long range effects (as shown in Figure 104).

Step IV – Compensate with Dose and/or Geometry the Target Layout: The final step of the proposed flow consists in checking the impact of the background dose over every shape present in the wafer/reticle. Hence, for every shape it is identified the impact on the weighted density map and this extra dose is compensated (by either adjusting the original shape's geometry or reducing the dose sent directly to it).



(a) (b)

Figure 104. (a) Weighted density in 2D plot, (b) Weighted density in 3D plot

Although the presented model was a single Gaussian distribution, this approach may employ any other model structure. Moreover, the presented strategy can be performed off-line (in contrast to machine correction). The advantage of employing off-line compensation relies on allowing a compensation using dose or geometry modulation. This strategy also has the advantage of being completely decoupled from other tool's internal corrections (like FEC). A result of the employing such flow is presented in section 7.3 in the results chapter of this thesis.

Chapter VII. Results

"However beautiful the strategy, you should occasionally look at the results."

- *Winston Churchill*

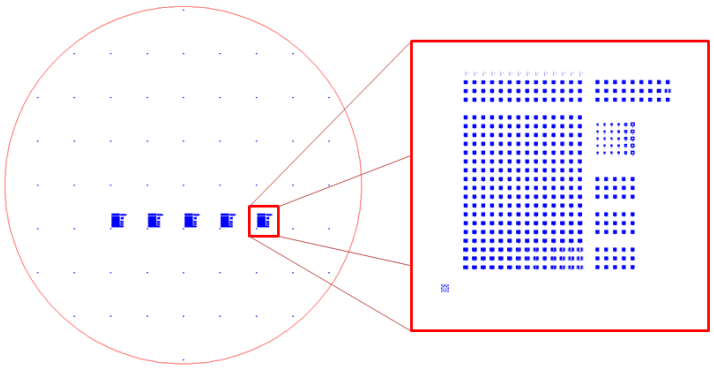
7. Results

7.1. Impact of Process Variability over Model Calibration

For e-beam lithography, the calibration of Compact Models is usually performed based on one or more cycles of pattern exposure and SEM measurements. Both the exposure procedure and the metrology one are subject to variability, which may induce inaccuracies on the calibrated model. Therefore, the goal of this example is to quantify such variability and evaluate its impact on model calibration [FIGUEIRO, 2013d].

7.1.1. Experimental Variability Quantification

In order to quantify and characterize variability, the following experimental test was performed. A test layout composed of a varied sort of patterns was exposed five times on a wafer, as presented in Figure 105.



(a) (b)

Figure 105. (a) The map of the test wafers and (b) detail of one of the five identical instances of the test layout.

The layout was composed of several different structures, varying in critical dimension (CD) from 40nm to 1µm and pattern density from completely isolated to extremely dense (Line Between Pads (LBP) and inverted Isolated Line (iISO)) passing by several L/S blocks (LSB) of different densities. The pattern types are presented in Figure 106: (a) L/S Block (LSB), (b) Line End Blocks (EOL), (c) Isolated Line (ISO), (d) Line Between Pads (LBP) and (e) Inverted Isolated Line (iISO).

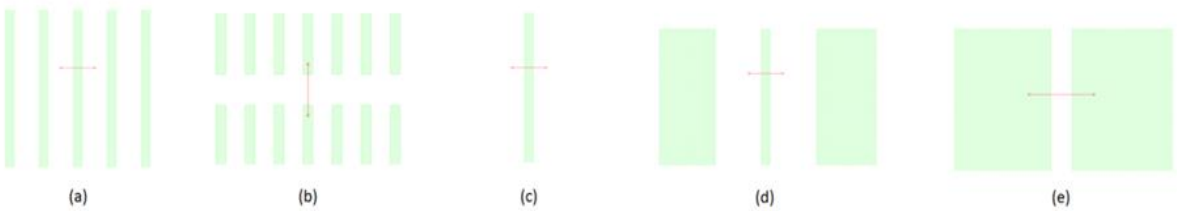


Figure 106. Examples of structures present on the layout: (a) LSB; (b) EOL; (c) ISO; (d) LBP; (e) iISO

The proposed test layout was exposed using a Vistec SB3050DW shape-beam tool at 50keV. The 300 mm wafer was coated with a commercially available positive chemically amplified e-beam resist. After the layout exposure, 350 points per chip were selected for measurement. The points were exactly the same for all five chip instances. The metrology was performed using two different and unmatched CD-SEMs, a Hitachi CG 4000 using an acceleration voltage of 0.5kV and an Applied Materials Verity 4i, with acceleration voltage of 0.8 kV. The tools are mentioned as Tool A and Tool B since the purpose here is not to compare the tools but to present different sources of variability. The measurements were performed over the same patterns on the same wafer. The metrology points were slightly shifted from the Tool A measurements to the Tool B measurement in order to prevent charging or resist shrinkage over the patterns. The total number of measurements during this experiment was 3500. After the measurements, the outliers coming from both measurements were filtered and, after that, the mean standard deviation of the same pattern for each tool was calculated. Finally, the mean value for each test pattern, for each tool, was calculated and the bias between both tools was calculated. Figure 107 shows a flow of the experimental test.

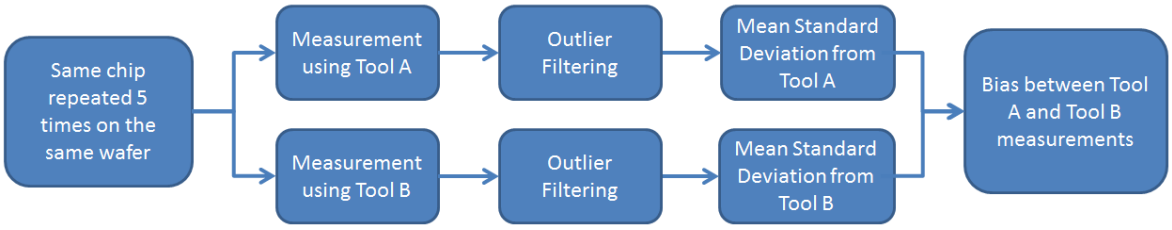
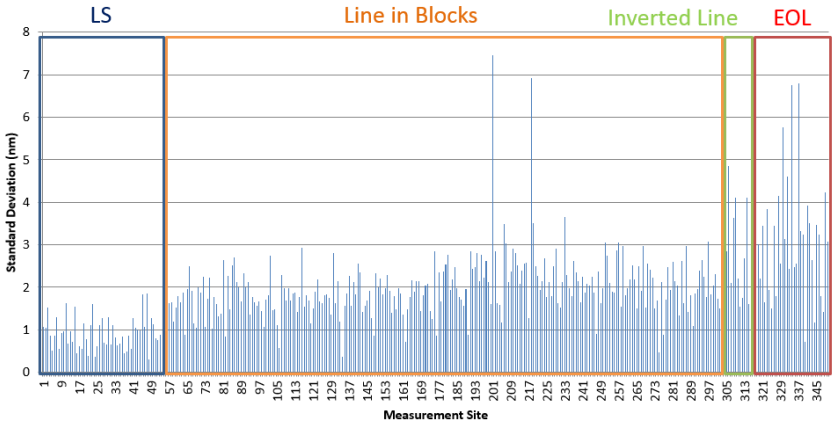


Figure 107. Flowchart illustrating the experimental test.

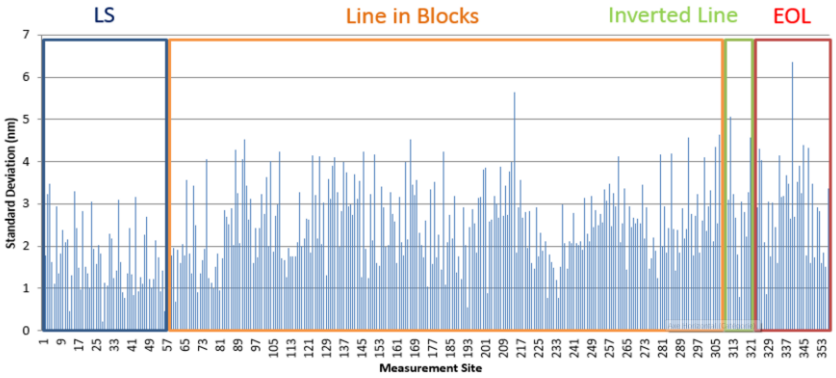
7.1.1.1. Standard Deviation for Tools A and B

The evaluation of the standard deviation of a given CD-SEM over the test patterns indicates the repeatability of the fabrication process and of the metrology. Since the PEC is performed using a model that does not take into consideration statistical variability, the standard deviation of a process is a metric of the best result a deterministic model may attain. Figure 108 (a) presents the standard deviation calculated for the 350 evaluated points for Tool A while Figure 108 (b) shows the same values for Tool B. The results are also summarized in the histograms presented in Figure 109 (a) for Tool A and Figure 109 (b) for Tool B and by Table 24. From the histograms presented in Figure 108 (a) and (b) it is possible to remark that the deviation is more or less uniformly distributed among the test patterns.

Both tools present mean standard deviations of around 2nm to 2.5nm, which is quite significant considering that in most cases calibration won't be performed over the mean value of several measurements over the same test pattern, but in only one instance of it. Moreover, the maximum standard deviation values reach 7.45nm and 6.36nm. Notice that these standard deviations are evaluated after outlier filtering.



(a)

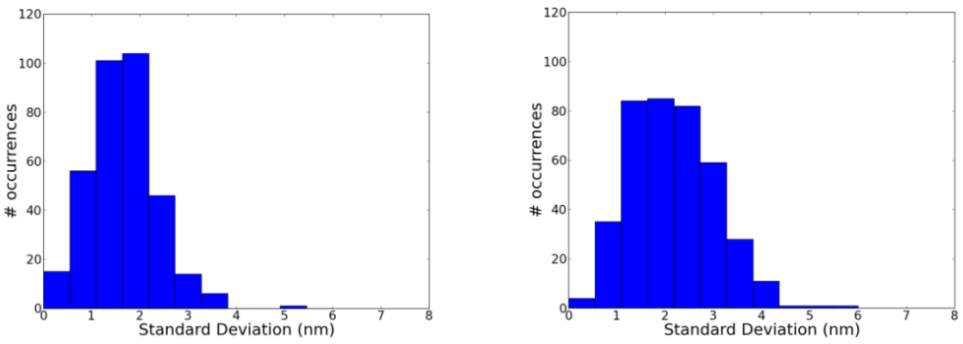


(b)

Figure 108. Standard deviation values per test pattern on the test layout measured by (a) Tool A and by (b) Tool B

Table 24. Mean and maximum standard deviation value over the entire set of patterns for Tools A and B.

	Tool A	Tool B
Mean Standard Deviation	1.98 nm	2.47 nm
Maximum Standard Deviation	7.45 nm	6.36 nm



(a) (b)

Figure 109. Histograms of the standard deviation values for the test patterns measured by (a) Tool A and (b) Tool B.

Moreover, it was possible to observe a significant difference in the standard deviation depending on the type of the pattern measured. Table 25 presents the standard deviation value evaluated by pattern type for both tools. . Observe that CD measurements of isolated lines and of L/S grating patterns are stable, presenting mean standard deviation inferior to 2nm and maximum standard deviation inferior to 3.5nm. On the other hand, measurements of opaque features (Isolated Inverted Line and End-of-line) and of line between pads presented less stability, with mean standard deviation going up to 3nm and maximum standard deviation as high as 7.45nm.

Table 25. Mean and maximum standard deviation values for Tools A and B by pattern type.

Pattern Type	Tool A		Tool B	
	Mean Standard Deviation	Maximum Standard Deviation	Mean Standard Deviation	Maximum Standard Deviation
ISO	0.62 nm	0.98 nm	1.83 nm	3.48 nm
iISO	2.72 nm	4.24 nm	2.42 nm	3.47 nm
LSB	1.04 nm	1.86 nm	1.67 nm	3.29 nm
LBP	1.97 nm	7.45 nm	2.5 nm	5.64 nm
EOL	2.74 nm	6.8 nm	3.08 nm	6.36 nm

7.1.1.2. Measurement bias between Tools A and B

Since the two CD-SEMs were not matched, it is expected to find a difference between the measurements obtained by each one. Therefore, an interesting evaluation is to compute the difference between the mean values for each test pattern. Figure 110 shows the difference per test pattern for Tool A minus Tool B mean measurement values. Notice that test patterns presented from position 300 to 350 (iISO and EOL) the measurements correspond to the distance between exposed patterns instead of patterns widths. For this reason, one may observe that there the bias between the two tools change of behavior. For instance, if a tool is consistently presenting higher measurement values for isolated line, it will present lower measurement values for an isolated inverted line.

In this way, it is possible to observe that there is a significant bias from the mean results from both CD-SEMs. Moreover, this bias appears to be systematic. If we compensate (inverse the sign) of the opaque patterns, the maximum absolute difference is 34.88nm, while the absolute mean difference is 10.93nm and standard deviation is 5.38nm. If one consider that the measurements were performed over the exactly same patterns, a difference of almost 11nm is significant. Even if one divide it among the two tools, the bias factor of the metrology would be around 5.5nm. A histogram of this bias per test pattern is presented in Figure 111. It is important to remark, however, that the bias was little impacted by the pattern type, as all pattern types presented similar absolute mean measurements and similar standard deviations.

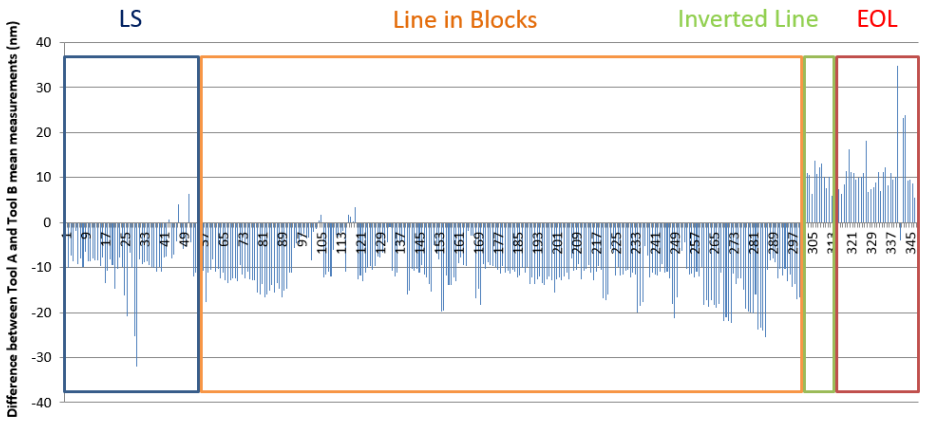


Figure 110. Difference between the means of the measurements performed by each of the two CD-SEM. Patterns from measurement site 300 to 350 are opaque.

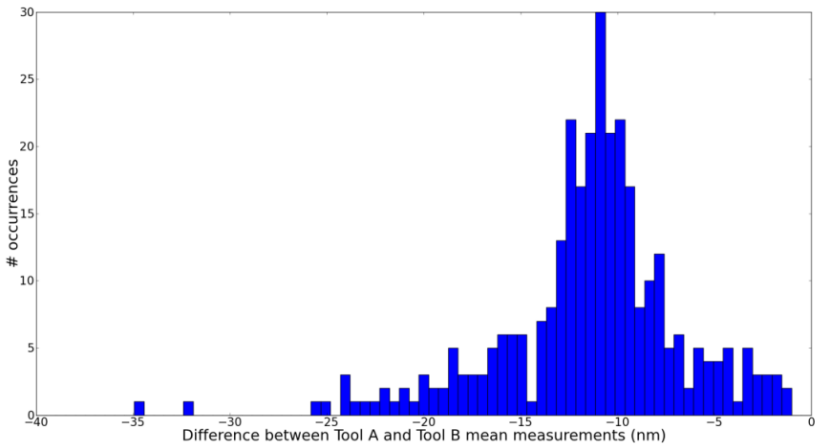


Figure 111. Histogram showing the difference between the means of the measurements performed by each CD-SEM using a tone aware evaluation

7.1.2. Variability Impact over Model Calibration

In order to verify the impact of metrology variability over the model calibration procedure, a set of calibration tests were performed. The main procedure is to employ a state-of-the-art model calibration algorithm [FIGUEIRO, 2012b] and using as inputs the same test layout used in the experimental exposure tests. The input data used was synthetically generated using simulation data coming from the simulation of triple-Gaussians PSFs. This model presents five parameters to be optimized. The range used for each parameter is presented in Table 26.

The synthetic data is then modified by adding some variability, as explained later in Section 7.1.2.1. All the results presented in this section come from the mean error value extracted from a procedure repeated five times. The error values presented are the mean relative errors in relation to the original model parameter values.

Table 26. The range used for each model parameter

Parameter	Minimum Value	Maximum Value
α	15	30
β_1	300	2000
η_1	0.1	0.25
β_2	5000	10000
η_2	0.2	0.45

7.1.2.1. Variability Generation

In order to create the variability on the original data, a normal distribution of noise was considered. Equation (95) shows the standard normal distribution formula used.

$$\frac{1}{\pi\sigma^2}\exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) \quad (95)$$

Where μ is the mean value, or in this specific test, the metrology bias or systematic error. σ is the standard deviation of the variability, the metrology random error parameter.

7.1.2.2. Random Error Only

Random error only test consists of adding a variability to simulated data over the calibration layout (by using equation (95) to generate an error factor for each simulated value). The variability range evaluated was from 3nm (the average variability observed in the previous test) to 10nm (above the maximum variability observed). The resulting simulation + random error is used as measurements in a calibration procedure, which is repeated five times. The results corresponding to each sigma variability (σ) evaluated are presented in Table 27. The first line is a reference without any error in the synthetic data, showing that the calibration algorithm was able to find the correct solution in all five runs and that the test layout presents enough sensitivity for determining each parameter exact value.

Table 27. The parameter estimation error according to the variability sigma used to inject error in the synthetic data

Variability Sigma	α Error	β_1 Error	η_1 Error	β_2 Error	η_2 Error
0 nm	0.00%	0.00%	0.00%	0,00%	0,00%
3 nm	1.42%	2.85%	6.00%	3.91%	0.58%
6 nm	5.71%	10.07%	14.63%	5.93%	1.58%
10 nm	6.23%	26.80%	18.25%	19.41%	4.93%

Up to 3nm sigma variability, errors remain below 6% so results are robust and accurate PSF determination is still possible. This robustness is a consequence of the high number (350) of metrology measurements and their sensitivity to the parameters of the model

used. For sigma variability of 6 nm and 10 nm, some of the parameters value errors were above 10 %. This analysis indicates the level of precision one may expect of the determination of each parameter under a certain level of variability of the process.

Table 28 presents the mean MSE (Mean Square Error) values from the five calibration procedures for each variability evaluated. The value of the "induced" MSE from the inserted variability is also presented. This is the MSE value that obtaining exactly the same model would produce. For a normally distributed variability, the "induced" MSE is exactly the variance of the normal distribution. Notice that the resulting MSE is always very close to the induced one. This indicates that the calibration procedure is still able to minimize the MSE (the actual calibration error may be evaluated by the difference between the MSE and the induced one). Moreover, this is a sign of a robust fit since the model does not absorb the injected error. This effect is observed when the resulting MSE is smaller than the "induced" one, indicating a fit of the noise added (overfit).

Table 28. The MSE values resulting from the model calibration and the induced ones

Variability Sigma	MSE (nm²)	Induced MSE (nm²)
3 nm	9.33	9.00
6 nm	36.14	36.00
10 nm	112.24	100.00

7.1.2.3. Systematic Error Only

The test for the systematic error only consist in adding a simple bias to the simulated values of all exposed patterns. Table 29 presents the parameter estimation error according to the bias value used. The range of bias tested were from 1nm to 5nm because that represents realistic values for general process biases.

In this case the degradation of the quality of parameter values is more significant than for the random error only test, especially for α and η_2. This may be explained by the fact that α value is the one that directly impact the width of all shapes of the test pattern, as well as η_2 is the one that dominates the ratio between α impact and other PSF components impact. The couple β_1 and η_1 comes right before, but the bias impact over their value is a little less significant because some patterns do not suffer from their direct influence (mainly isolated patterns or patterns presenting a clear region around it of the order of β_1).

Another important aspect is the analysis of the MSE results, presented in Table 30. Notice that the MSE value is always small, indicating that the calibrated model has absorbed the impact of the injected bias. Although it may look as a good side-effect, it is in fact a tricky one since it indicates that the model is incorporating a bias, which is not part of its structure. This overfitting behavior leads to bad predictions when it comes to simulate other patterns not present in the calibration patterns set. This lack of predictivity may be solved by including the bias in the model and add patterns to the calibration layout that allow this parameter to be properly optimized.

Table 29.The parameter estimation error according to the bias used to inject error in the synthetic data

Bias	α Error	β_1 Error	η_1 Error	β_2 Error	η_2 Error
0 nm	0.00%	0.00%	0.00%	0,00%	0,00%
1 nm	2,17%	2,50%	0,27%	0,19%	2,50%
2 nm	5,34%	4,08%	2,63%	0,22%	4,65%
3 nm	7,08%	6,70%	2,78%	0,13%	7,04%
5 nm	11,82%	9,19%	6,40%	1,98%	11,38%

Table 30. The MSE values resulting from the model calibration

Bias	MSE (nm^2)
1 nm	0.027
2 nm	0.098
3 nm	0.173
5 nm	0.399

7.1.2.4. Systematic Error and Random Error Combined

The final example is a calibration considering both random and systematic errors combined. In this case the mean results from five different tests indicate that both phenomena are not correlated and what is observed is almost a sum of both impacts (although η_1 and β_2 values does not respect this behavior, it might be due to the correlated sensitivity of these two parameters). The results are presented in Table 31.

Table 31. The parameter estimation error for the combined test of variability sigma and bias errors combined

Variability Sigma	Bias	α Error	β_1 Error	η_1 Error	β_2 Error	η_2 Error
3 nm	2 nm	6,31%	7,46%	6,85%	2,11%	6,76%

The resulting MSE of this test was $9.15nm^2$, which is very close to the "induced" MSE ($9\ nm^2$) coming from the introduced variability. This is usually within the acceptable range of precision for the calibration of a model, especially when performed over such level of variability.

7.1.3. Summary

The first part of this experiment presents experimental CD measurements of the same patterns when using different CD-SEM tools. The measurements from these two different tools presented a mean variance of 2.0nm and 2.5nm, respectively. However, for some specific patterns the variance could reach around 3.0nm, with some points going up to 7.45nm. Moreover, comparison between the mean values of the measurements coming from two CD-SEMs presented a significant metrology bias of around 5.5nm (estimated from the 11.0nm bias between processes). The bias itself is not an issue in the metrology perspective, since the machines are in different cleanrooms and are not matched at all. However, it reinforces the idea

that whenever modeling a process, the metrology step must be seen as part of the process being modeled.

The second part of this experiment evaluated the impact of random variability and systematic variability in the data used for model calibration. In the analysis of random error only, it is possible to observe that the set of calibration presenting 350 patterns was effective to cope with up to 3nm variability. The MSE values obtained indicate that the model calibration was robust, preventing from the overfitting of handling random error as something systematic (that should be present on the model). Systematic errors presented large impact in the parameter values. The model has no explicit bias parameter and the fact that the MSE continues to be small indicates an overfit (lack of predictivity). The solution for this issue would be to include a bias parameter to the model and provide a calibration set that could present sensitivity to it.

From the analysis of the values of MSE after model calibration in the presence of random and systematic variability it is demonstrated that a validation step after model calibration is mandatory since the MSE information alone may not assure the robustness of the calibrated model (as shown by lack of predictivity in the systematic variability experiment).

7.2. Model Calibration Use Cases

In this section some model calibration use cases are performed in order to demonstrate the capabilities of the model structures proposed in this work such as the Gamma PSF, Pearson PSF, Spline PSF and constant and variable resist threshold models. Three different model calibration use cases are presented. The first one consists in calibrating a Gamma PSF, a Pearson PSF and a Spline PSF based on a First Principle model and compare the results with traditional Gaussian PSF models. The second exercise is a calibration of experimental data performed over an EUV mask exposure where we were able to propose a calibration layout and validate the generate models with another exposure. Finally, the third exercise consist of calibrating a model over a set of data already available. The most significant results from this last exercise is the comparison of the 2D simulation result with SEM images of the patterns which demonstrated the ability of the model to predict when the exposure would no longer print as intended.

7.2.1. Gamma PSF

One of the strategies to calibrate a Compact model for e-beam lithography is to rely on First Principle model simulation. In order to demonstrate this use case a study to evaluate the ability of describing the energy distribution over a EUV stack is performed. A PSF model based only on Gaussian distribution is compared with one mixing both Gaussian and Gamma distributions, which is developed during this thesis work (shown in section 4.3.1.2) and first presented in [FIGUEIRO, 2012a]. The reference to compare comes from Monte Carlo simulations performed using the CASINO software [DROUIN, 2007]. The conditions of this simulation are: 5 million electrons simulated, acceleration voltage of 50keV, e-beam radius of 20nm and no tilt of the sample. The layer stack composition is 300nm of PMMA, 90nm of Ta over a SiO_2 stack. The MoSi multilayer has little impact over the energy distribution [TANABE, 2010].

The model calibration strategy used was Efficient Global Optimization (EGO) [JONES, 1998] and the cost functions used were MSE or MSLE (as described in section 2.2.1). The MSE cost function considers the square value of the error, therefore, it is interesting to evaluate if the higher energy portion of the PSF is properly fit. The MSLE cost function is interesting because it considers the fit in a global manner, giving equal importance to all portions of the curve.

The Gamma PSF was calibrated and, for comparison purposes, three different Gaussian PSFs were also calibrated, presenting either 3, 4 or 5 Gaussian distributions. The fitting results for the two cost functions evaluated in all four models are presented in Table 32.

Table 32. Results after fitting procedure over the 4 models evaluated with two different cost functions.

PSF Model	No. Of Parameters	No. Of Components	MSE	MSLE
3 Gaussians PSF	5	3	1.763 e^{-7}	1.436
4 Gaussians PSF	7	4	1.755 e^{-7}	1.364
5 Gaussians PSF	9	5	1.696 e^{-7}	1.338
Gamma PSF	9	4	1.681 e^{-7}	1.307

Figure 112 shows the simulation results for the EUV stack with the fitting results for (a) the 3 Gaussian PSF model, (b) the 4 Gaussian PSF model, (c) the 5 Gaussian PSF model and (d) the Gamma PSF model. Notice that, although the visual result for the fit of the 5 Gaussians PSF (Figure 112(c)) may look better when compared to the Gamma PSF result (Figure 112(d)), the Gamma PSF presents a better result both when fitting for a MSE and for the MSLE.

The simulation effort is usually proportional to the number of components of a model. This is due to the fact that each component is convoluted to the patterns. In order to evaluate if the proposed Gamma PSF is comparable to the 4 Gaussian PSF, a comparative performance test was performed. Both models were used in the correction for 7 designs with different sizes, number of shapes and complexities. The results are presented in Table 33, normalized by the execution time for the 4 Gaussian PSF model.

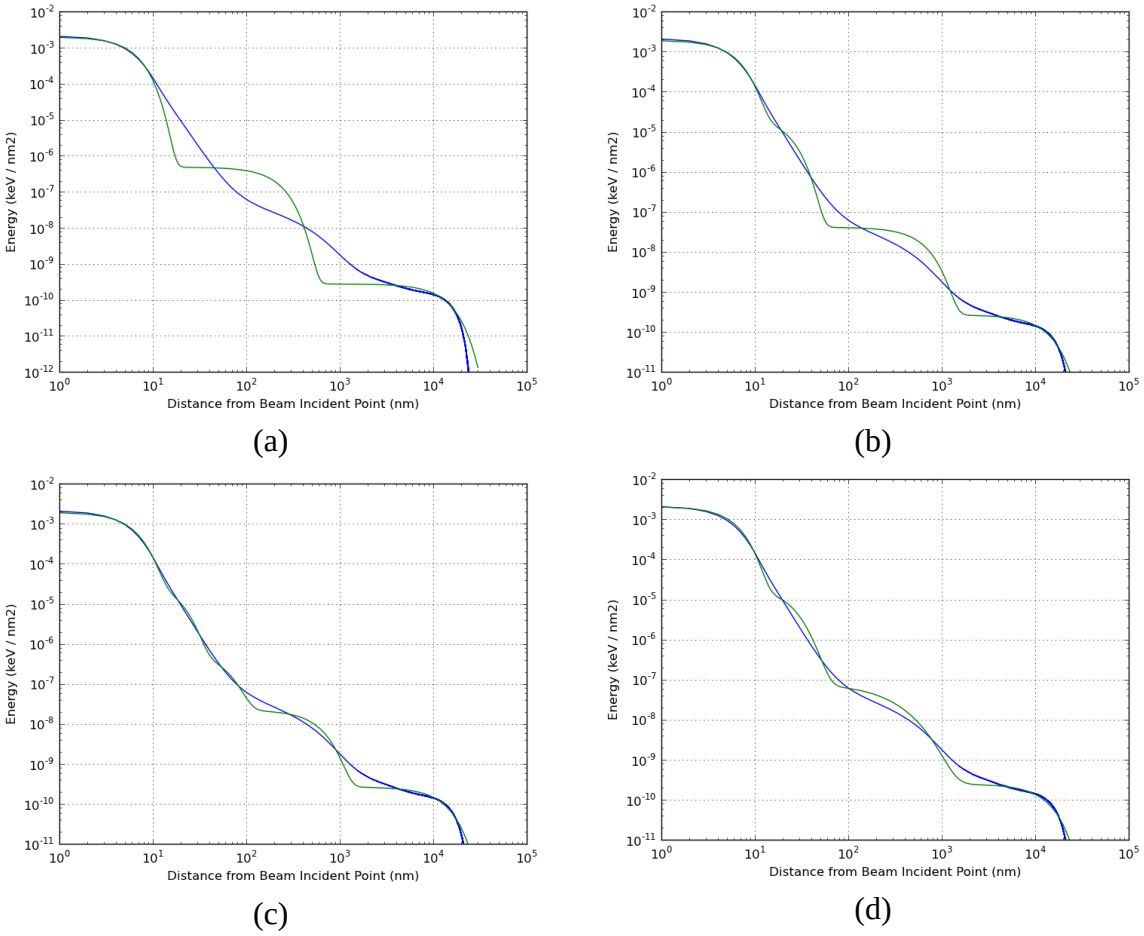


Figure 112. Fitting result for the (a) 3 Gaussian PSF, (b) 4 Gaussians PSF, (c) 5 Gaussians PSF and for the (d) Gamma PSF.

Table 33. Results after fitting procedure over the 4 models evaluated with two different cost functions.

	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 7	Mean
Gamma PSF	1.044	1.016	1.064	1.021	1.040	1.005	1.041	1.033

Notice that the Gamma PSF model degrades the performance of the correction algorithm in only around 3.3% while improving the model precision by around 4.2%.

7.2.2. EUV Mask Writing Process Calibration

An e-beam writing process for EUV masks installed in an industrial production environment required a model for performing PEC. One of the particularities of the EUV process (as discussed in section 1.2.1) is the different mask structure which impacts the electron scattering. Moreover, the precision requirements were so tight that no simple model could be employed. On the other hand, performance constraints were also relevant and using a model too complex was also out of question. For this reason, the best alternative was to first perform calibration using Pearson PSFs, which present the behavior similar to what is observed in EUV First Principle model simulations and present the characteristics that enable easier calibration (as discussed in section 4.3.1.5).

The calibration layout was composed of around 400 patterns, exposed at 4 different doses (10% underexposure, at nominal dose and 10% and 20% overexposure). The different doses are used to increase the robustness of the model to dose variation. The 400 patterns are the following: 50 LSB (CDs from 40nm to 1000nm and three different densities – 50%, 25% and 15%); 250 LBP (three different CDs: 100nm, 300nm and 600nm, and spaces varying from 60nm to 50000nm); 15 ISO (CDs from 40nm to 1000nm); 45 EOL (CDs from 80nm to 150nm, line widths from 60nm to 200nm and three densities – 50%, 25% and 15%); 10 iISO patterns (CDs from 80 nm to 750nm) and 30 LBP repeated from the first group for variation control purposed.

These patterns present different densities and, therefore, exposing such variety of patterns with a single exposure dose is not feasible. For this purpose, a pre-correction with a oversimplified model was performed. Basically, it attributes 4 different doses and set them in a way that very dense patterns receive less dose while the more isolated ones receive less. The final result for a single layout is shown in Figure 113.

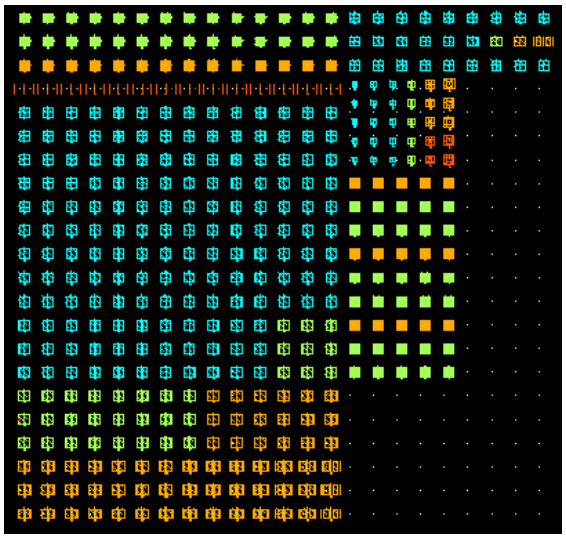
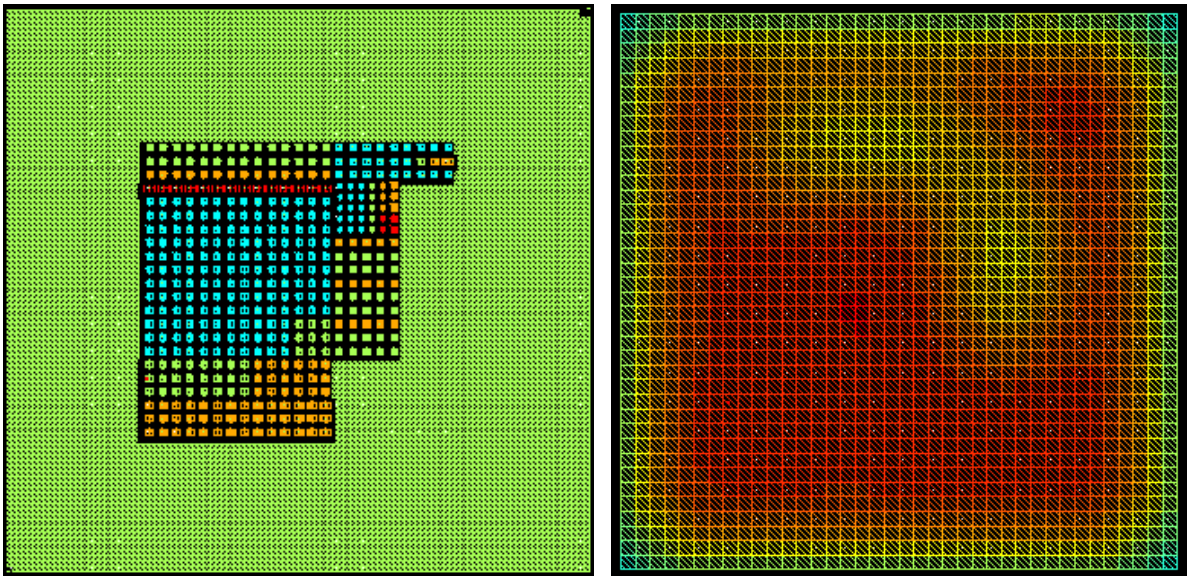


Figure 113. Floorplan of the calibration layout.

Moreover, since extreme-long-range effects were not to be handle at this point, we used a simple strategy of adding dummy contacts between the patterns in order to make the local density more uniform. Original densities ranging from 0% to 21% were reduced to a range from 10% to 16%. The general floorplan of the layout after adding the dummy patterns is shown in Figure 114.

The calibration procedure was performed based on the 1600 measurements available. Three different models were generated, one based on 5 Gaussian PSF and two based on a 4 Pearson PSF. The Gaussian PSF and one of the Pearson PSFs presented a constant threshold calibrated at the same time as the other parameters of the model, the other Pearson PSF replaced the obtained threshold by a surface model based on three metrics: CD Kernel, Space Kernel and Density Kernel. The result of the calibration for the Gaussian PSF presented an error RMS superior to 5nm while the Pearson PSF models presented RMS values of 2.91nm for the constant threshold and of 1.89nm when relying on the surface model.



(a) (b)

Figure 114. (a) Floorplan of the calibration layout after adding dummies and (b) the estimated extreme-long-range effect over the pattern due to density –values inside the patterns value from 10% to 16%.

As the results for the Gaussian PSF model were inferior, it was abandoned for the sequence of the evaluation. In order to verify the two Pearson PSF models, a new test layout was created. The test layout consisted of patterns present on the calibration layout and some others that present intermediate dimensions. The generated layout was composed of 106 patterns, distributed as shown in Figure 115. The dimensions of each point shown varied from one pattern to the next, with CDs varying from 80nm to 300nm, while spaces varied from 80nm to 2000nm and LSB had densities varying from 50% to 10%.

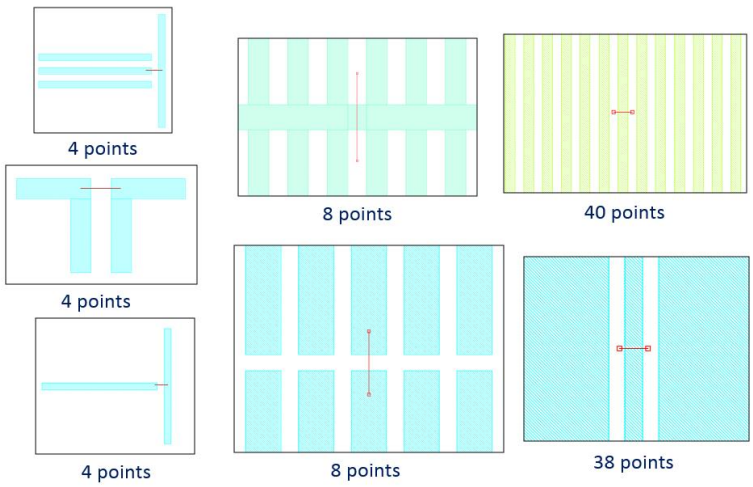


Figure 115. Exposed patterns for performing the validation of both models.

The proposed patterns were corrected using Inscale™ Dose and Geometry compensation [MANAKLI, 2007] relying on the prediction from both models. The two generated layouts were exposed and measured using exactly the same process from the calibration pattern.

7.2.2.1. *Pearson PSF + Constant Threshold*

The results obtained after correction using the Pearson PSF and the Constant Threshold strategy are represented in the histogram in Figure 116.

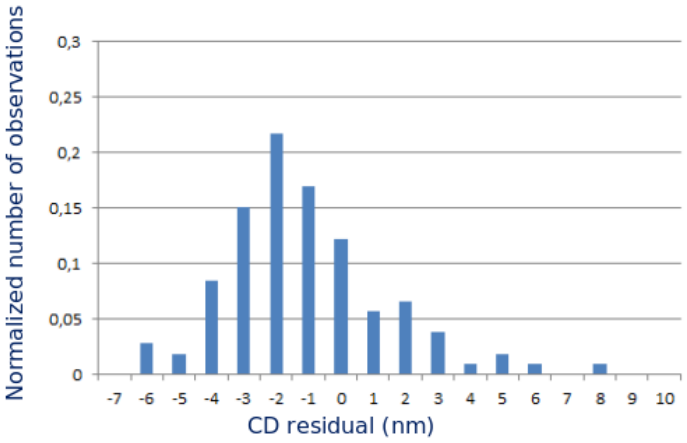


Figure 116. Histogram of the CD residual after correction based on the Constant Threshold Model.

Observe that the mode of the histogram distribution is not in zero, indicating a probable process shift in relation to the model and most likely to the previous exposure as well. However, the final result is still promising if one consider the RMS of 2.69nm and, moreover, the Standard Deviation of 2.47nm, which compensated for an eventual shift on the data and is, therefore, a better metric of the process distribution error. This value is quite good considering that the variability one may observe in this process is around 2nm. One may also evaluate the minimum value and the maximum values of the error, as well as the difference between the maximum absolute error and the average error, which would be the maximum absolute error after compensating for the average shift. All this information is provided for the entire set of patterns in Figure 117(a) and presented for the subset of small CDs (only CDs from 80nm to 100nm), shown in Figure 117(b). This second set presents only 15 patterns but is relevant since it represents the most critical features for this process resolution.

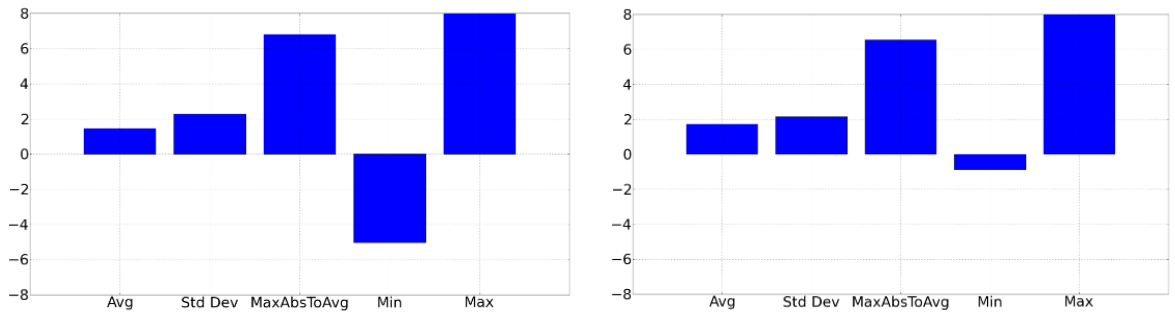


Figure 117. Average Error, Standard Deviation, Maximum Absolute Error To Average, Minimum Error and Maximum Error for (a) the entire set of patterns and for (b) only patterns with CDs between 80nm and 100nm..

7.2.2.2. *Pearson PSF + Variable Threshold*

The results obtained after correction using the Pearson PSF and the Variable Threshold strategy are represented in the histogram in Figure 118.

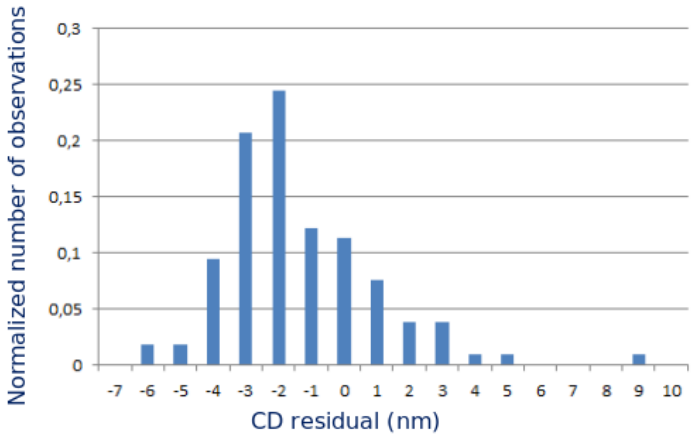


Figure 118. Histogram of the CD residual after correction based on the Variable Threshold Model.

Observe that the mode of the histogram distribution is not in zero, as observed with in the Constant Threshold model experiment. This reinforces the idea of a process shift. However, the final result is still promising if one consider the RMS of 2.65nm and, moreover, the Standard Deviation of 2.29nm, which compensated for an eventual shift on the data. These values represent a high quality model putting into perspective that the variation of this process is of around 2nm. Notice that although the RMS value of both the Constant and the Variable Threshold models are very similar, one may observe a small improvement by using the Variable Threshold when comparing the Standard Deviations (from 2.47nm to 2.29nm).

The same information provided to the Constant Threshold Model in Figure 118 is also presented for the Variable Threshold in Figure 119.

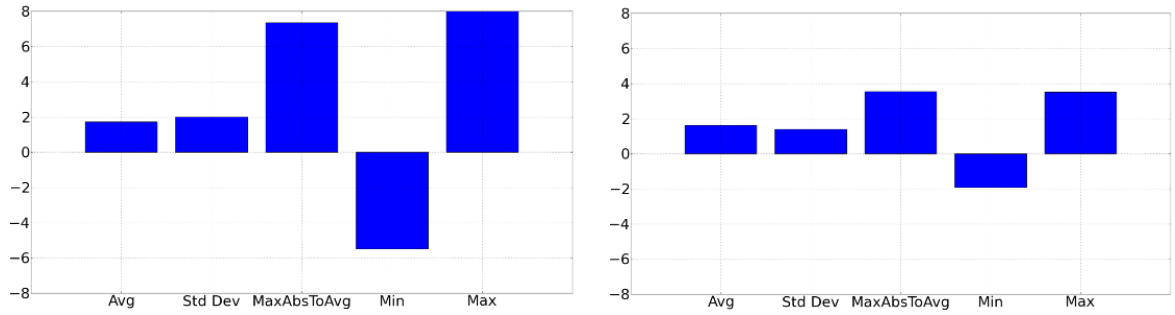


Figure 119. Average Error, Standard Deviation, Maximum Absolute Error To Average, Minimum Error and Maximum Error for (a) the entire set of patterns and for (b) only patterns with CDs between 80nm and 100nm.

It is interesting to remark that the Variable Threshold model, although presenting comparable results in the entire set of patterns, provide such a significant improvement for the critical ones (comparison between Figure 117(b) and Figure 119(b)). The errors range was reduced from around -5.5nm to 8nm to a range of around -2nm to 3.5nm. This is not properly reflected on the general model statistics since the proportion of such patterns in comparison to the total is not much representative (only about 14%).

With this exercise one may observe the capabilities on employing a model relying on the Pearson PSF, which provided good results for the pattern under calibration. Moreover, it was possible to certify the advantage of using the Variable Threshold Model for further improving the performance in critical dimensions, where a Constant Threshold Model fails to describe accurately.

7.2.3. Calibration Using Only Pre-available Data

In this particular case the calibration layout was not selected by using the proposed method in this paper, but was imposed by the customer. In a situation like this it was necessary to handle the process as a complete black box, relying only on the information provided by the layout and the metrology information. The calibration layout provided contained 737 patterns distributed in 8 different categories as shown in Figure 120. The dimensions for each pattern varied from 40nm to 2µm.

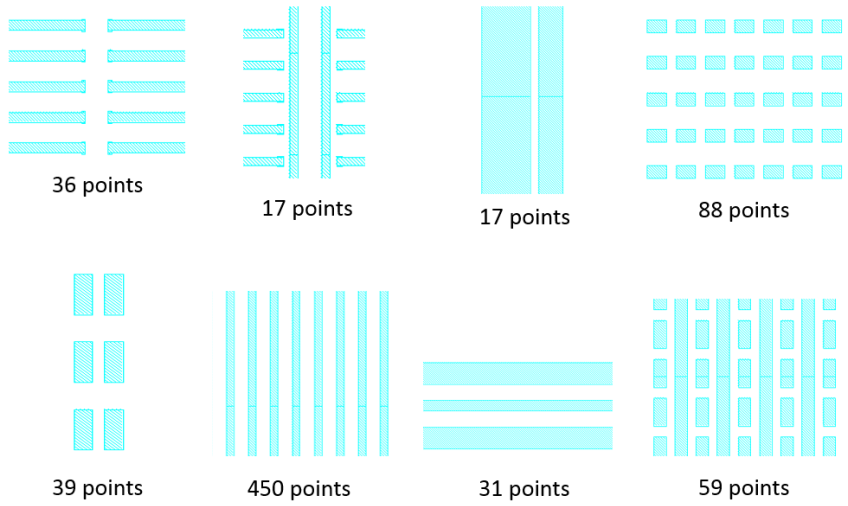
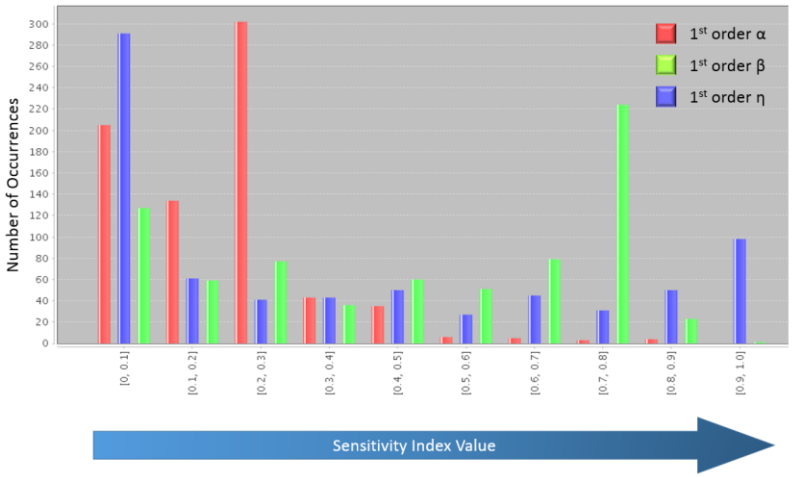
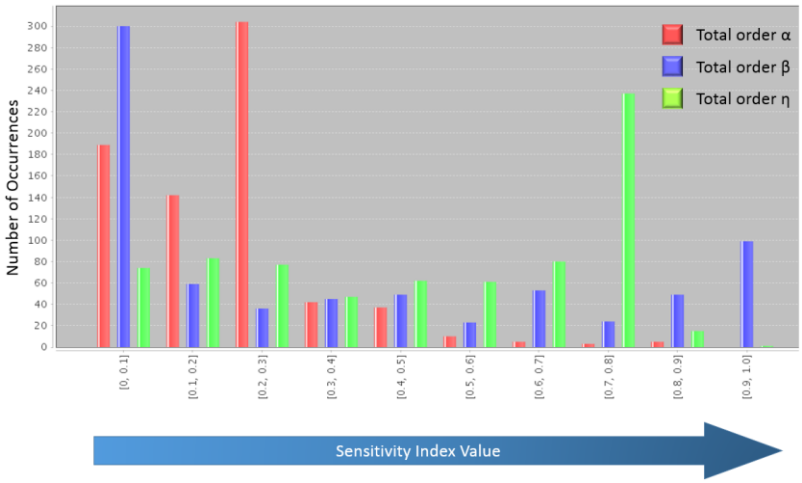


Figure 120. Different types of patterns and their number of variations on the calibration layout.

First, a sensitivity analysis was performed over the given patterns to identify the pertinence of calibrating more complex models. The first candidate is a simple 2 Gaussian PSF. The histogram indicating the number of patterns according to the sensitive index value of each model parameter is shown in Figure 121.



(a)

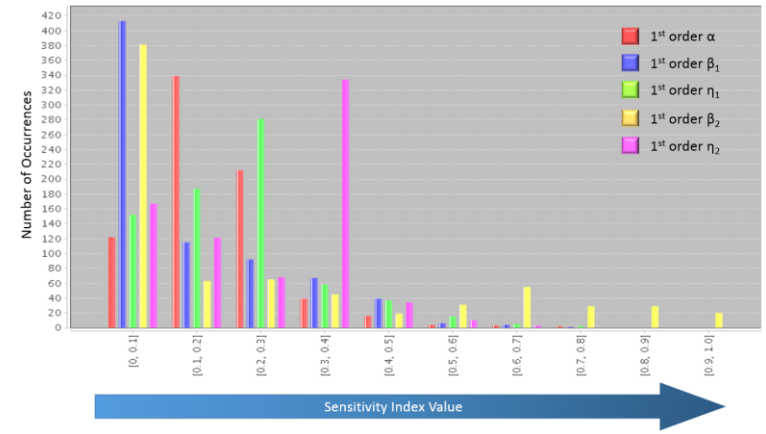


(b)

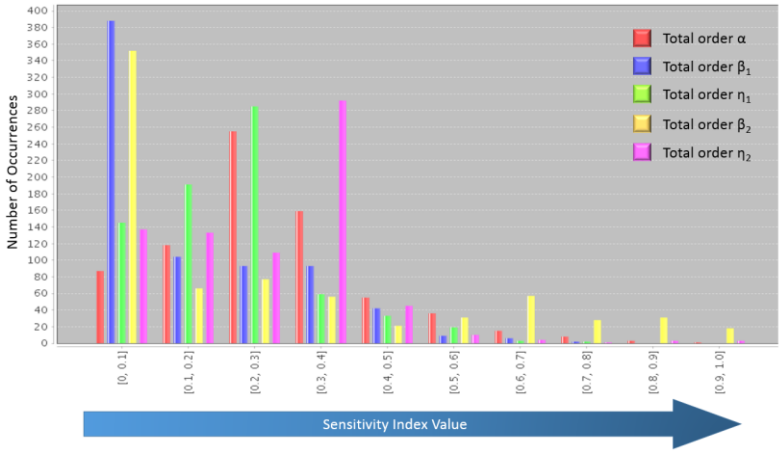
Figure 121. Histogram of calibration patterns for each 2 Gaussian PSF model parameter for (a) Sobol' 1st Order indices and (b) Total Order indices.

Observe that the sensitivity to the α parameter is not much significant (out of the 737 patterns, only around 20 present 1st order sensitivity indices above 0.5), as shown in (Figure 121(a)). However, there is a significant number of patterns that present high sensitivity to η_1 parameter (round 240 patterns with 1st Order sensitivity above 0.7) and around 200 patterns with 1st Order indices above 0.7 for the beta1 parameter.

The next evaluation considered the 3 Gaussian PSF model (presented in Figure 122) and the 2 Gaussian 1 Pearson model (presented in Figure 123).



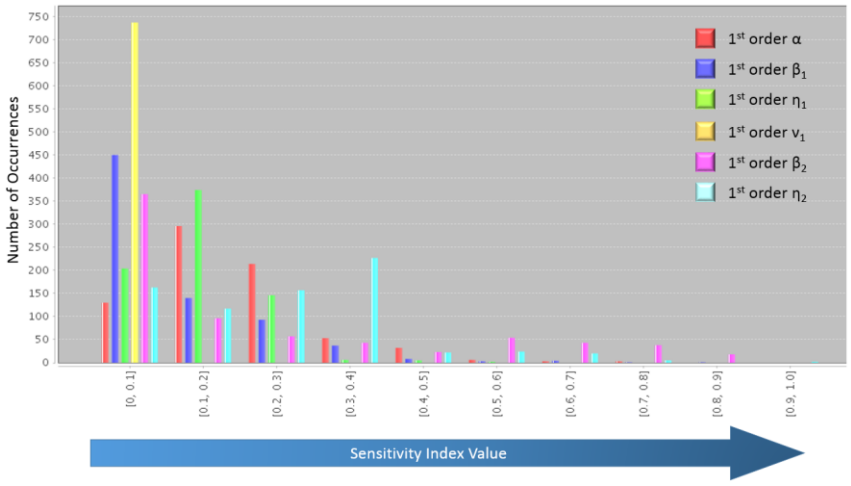
(a)



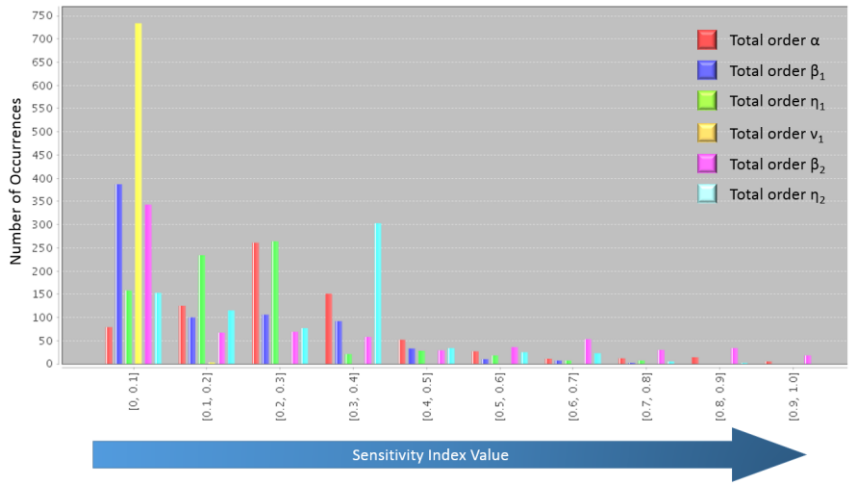
(b)

Figure 122. Histogram of calibration patterns for each 3 Gaussian PSF model parameter for (a) Sobol' 1st Order indices and (b) Total Order indices.

Notice that the only parameter presenting a high sensitivity is the β_2. The use of two distributions and the lack of patterns to separate the two ranges actually make the sensitivity to η_2 and η_1 to be low. This makes using this model not so interesting because one may tell that the parameter values will not be accurately determined.



(a)

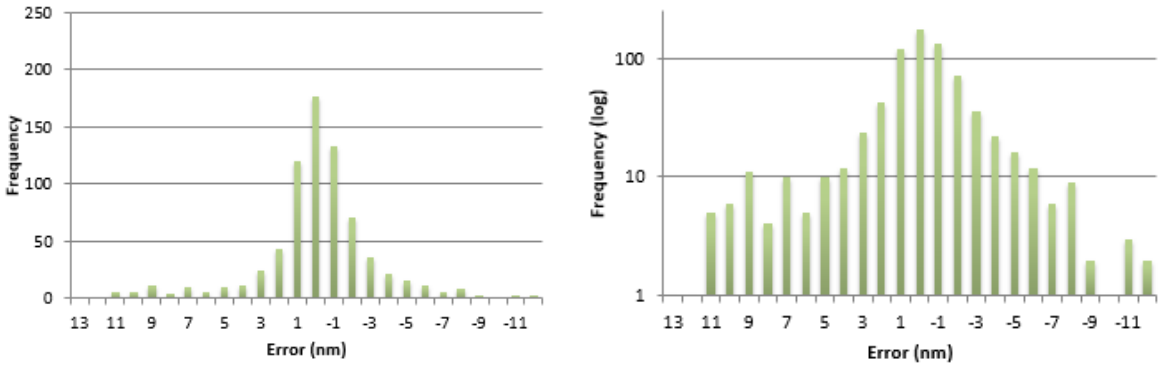


(b)

Figure 123. Histogram of calibration patterns for each 2 Gaussian 1 Pearson PSF model parameter for (a) Sobol' 1st Order indices and (b) Total Order indices.

The same lack of sensitivity from the 3 Gaussian PSF is observed for the 2 Gaussian 1 Pearson PSF. This is expected since the Pearson function may ultimately mimic a Gaussian behavior. However, one may further notice that there is virtually no pattern presenting any sensitivity to the ν parameter (which controls the shape of the Pearson distribution). This may help concluding that employing a Pearson distribution is not interesting since there is no actual information allowing its accurate calibration. Therefore, the best approach on this case is to perform the calibration using a 2 Gaussian PSF model, which the patterns present a more significant sensitivity, with the exception of the α parameter.

The exposed data was already corrected by the exposure tool algorithm. The objective of the model was to predict the edge placement error (EPE) after this correction. Comparing the experimental results with the intended target one may observe the error distribution as shown in Figure 124 (a) and in logarithmic scale in Figure 124 (b).

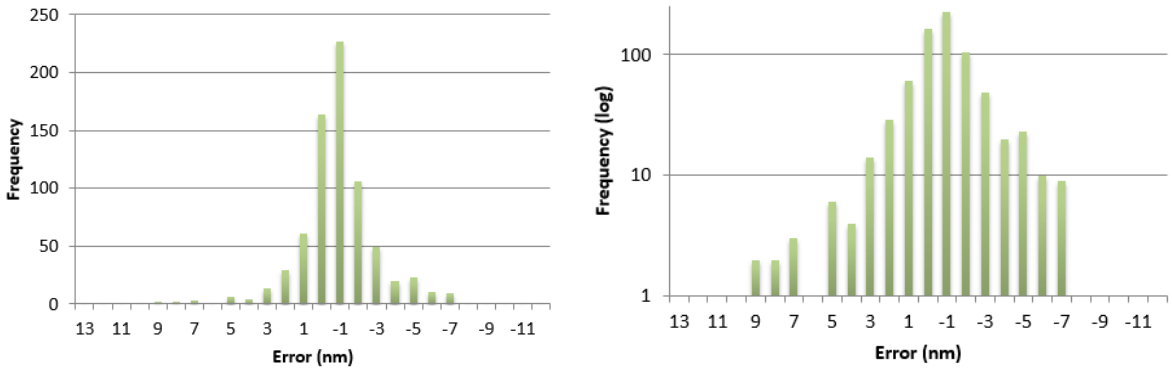


(a) (b)

Figure 124. Histogram of the frequency of errors in the process after tool PEC using unknown parameters presented in (a) normal scale and (b) logarithmic scale.

This is the exposure result with the PEC algorithm activated in the exposure. This means that this is the kind of error one observe in real masks after performing the state-of-the-art corrections available in commercial e-beam mask writers. The errors RMS is 3.42nm while the mean absolute error is 2.29nm and the maximum absolute error is 18.3nm. This maximum error is considerably high especially if one consider the errors in the circuit such difference may give rise. Moreover, in a stable process such as this one the variability is around 2nm, which if considerably lower than the 3.42nm presented in this distribution.

By providing this data to the calibration algorithm one may create a model to explain the residual error from the machine correction. The resulting model errors are shown in Figure 125.



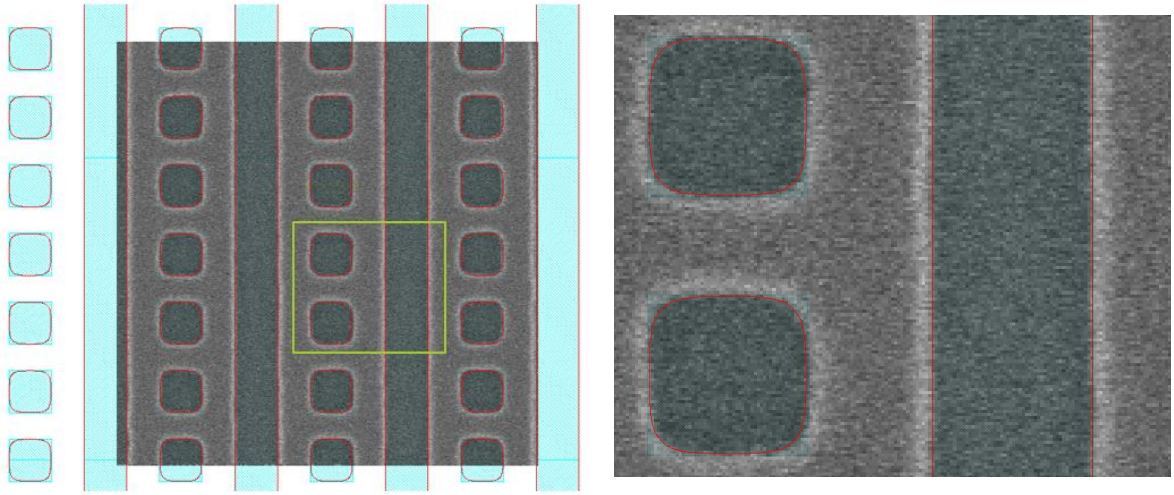
(a) (b)

Figure 125. Histogram of the frequency of errors in the process after performing a 2 Gaussian PSF modeling over the previous set of data presented in (a) normal scale and (b) logarithmic scale.

The RMS is 2.31nm while the mean absolute error is 1.49nm and the maximum absolute error is 9.5nm. Observe that now the RMS value is close to the variability one observe in this kind of process. Notice that such test already includes a metrology variability since the comparison is performed against experimental data.

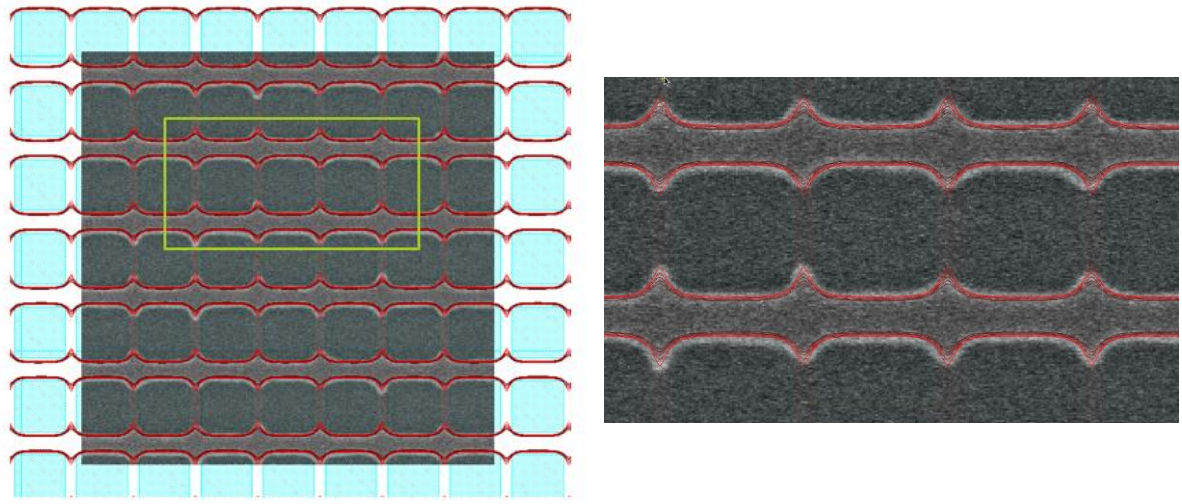
Another strategy used to evaluate the quality of the generated model was to perform 2D simulation contours and overlay them in the SEM images available. One example of such overlay is shown in Figure 126(a) and on detail on Figure 126(b). Moreover, even overlaying

the simulation with a SEM image that did not provided measurements due to bridging also indicated the good predictability of the model since the simulated contour also indicated the bridging in these areas (as shown in Figure 127(a) and in detail in Figure 127(b)).



(a) (b)

Figure 126. (a) Simulation contour using the calibrated model and compared with the overlay of the SEM image of the exposure. (b) Shows a detail of the simulation contour.



(a) (b)

Figure 127. (a) Simulation contour using the calibrated model and compared with the overlay of the SEM image of the exposure. (b) Shows a detail of the simulation contour detecting the bridging observed on the experiment.

The interesting result in this exercise is that even using a simple model, it is possible to significantly improve the correction results obtained by this exposure. One may assume that by relying on the prediction provided by the calibrated 2 Gaussian PSF a PEC algorithm would improve the RMSE by about 1.1nm while the mean absolute error by 0.8nm and the maximum error would drop from 18.3nm to 9.75nm. It was not possible to obtain even better prediction due to the limitations on the calibration layout used that did not provided enough sensitivity to

the mid-range, making not feasible to have the model covering this region. Finally, a model calibrated not regarding the SEM pictures provided good results on the contour determination and was able to identify cases such as the bridging detected in Figure 127.

7.3. Extreme-Long-Range Calibration Results

A set of test patterns was used in order to verify and calibrate the so called extreme long range effects in e-beam lithography.

First, a comparison between two exposed wafers is performed. The first wafer, presented in Figure 128(a)), contained the evaluation patterns but no dummy blocks to cause the density related extreme-long-range effect. It is used as a reference. The second wafer (shown in Figure 128(b)) contains a density dummy block representing a 50% density in a region 50mm wide that was placed 1mm away from the evaluation pattern set.

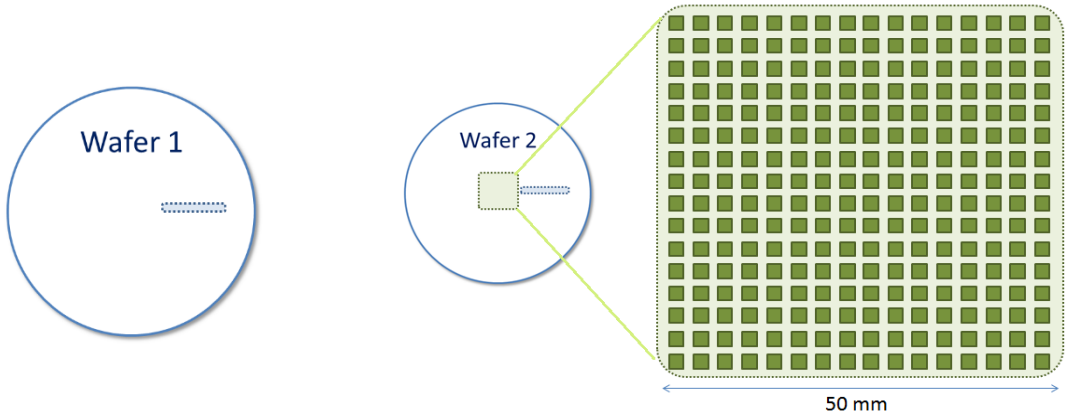


Figure 128. (a) Exposed wafer including only the test patterns and (b) exposed wafer including the 50 mm wide dummy blocks with 50% density on the center.

The measurement results of the evaluation pattern for both are presented in Figure 129, Figure 130 and Figure 131, for the isolated lines, line/space blocks and line between pads, respectively. They are represented as a function of the distance from the center of the wafer.

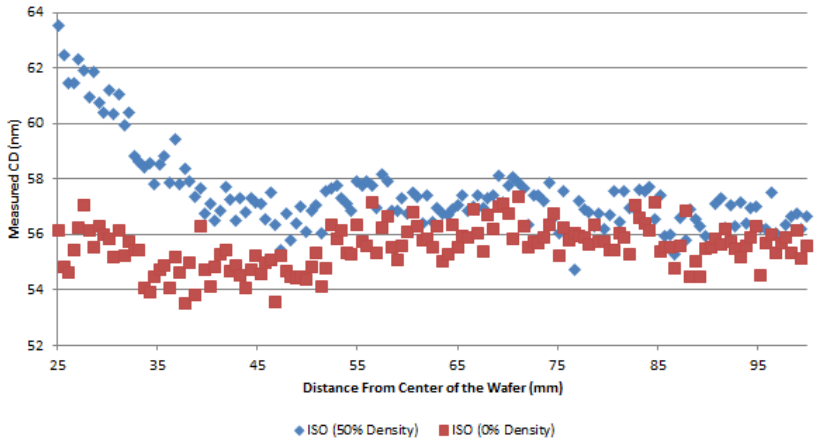


Figure 129. Metrology results for the isolated lines pattern (wafer 1, presenting no dummies in red and wafer 2, presenting 50% density dummies in blue).

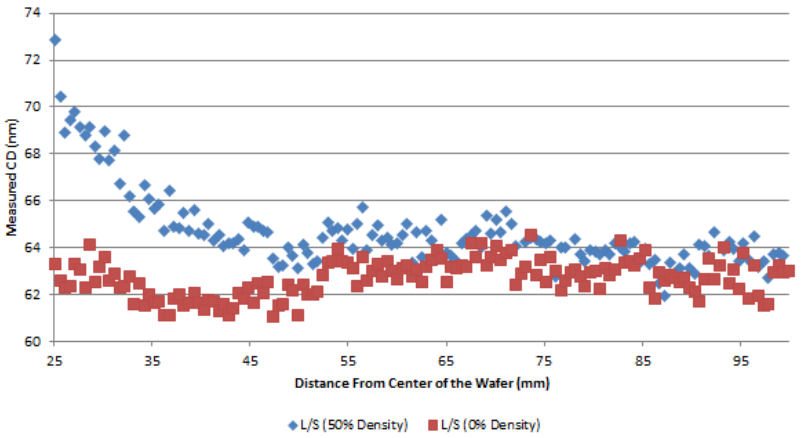


Figure 130. Metrology results for the line/space pattern (wafer 1, presenting no dummies in red and wafer 2, presenting 50% density dummies in blue).

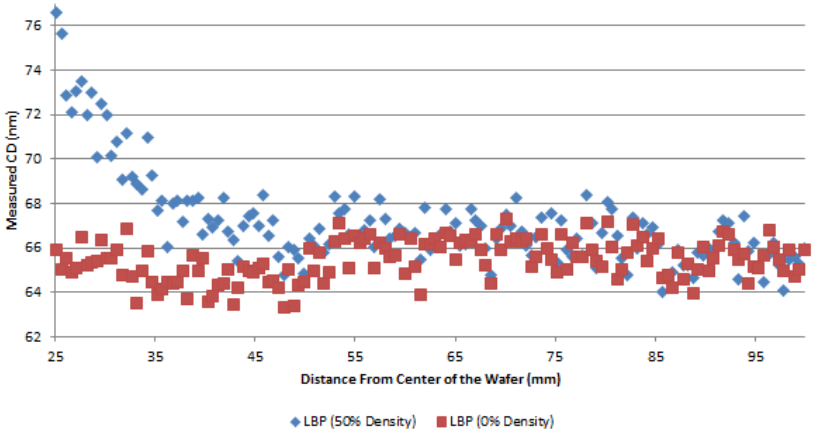


Figure 131. Metrology results for the line between blocks pattern (wafer 1, presenting no dummies in red and wafer 2, presenting 50% density dummies in blue).

As it is possible to remark, for all cases the dummy block impacted the patterns exposed up to 20 mm away and this impact was up to around 10nm (15%) in critical dimension.

The next step of this evaluation is to perform a calibration of a model to explain the extreme-long-range phenomena observed. However, in order to have a reference of the quality of the model, first it is necessary to understand the process variability by evaluating the RMS values of the same patterns printed several times without a dummy block. The resulting RMS is of 0.82nm (the histogram is presented in Figure 132). This is a quite low value and it is only achieved because of the use of a test pattern that is stable and for which the SEM was calibrated for.

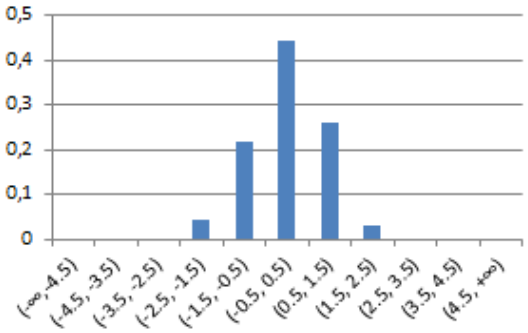


Figure 132. Histogram presenting the process variability evaluated over the measurements performed in the wafer with no dummy block.

The model of the extreme long range effect was calibrated as a single Gaussian distribution. This model was calibrated using wafers printed with four different dummy block densities: 0% (no dummies), 10%, 30% and 50%. The resulting calibration determined a distribution sigma of 4mm and presented an RMS value of 1.13nm (comparing measurements and simulations, as shown in Figure 133).

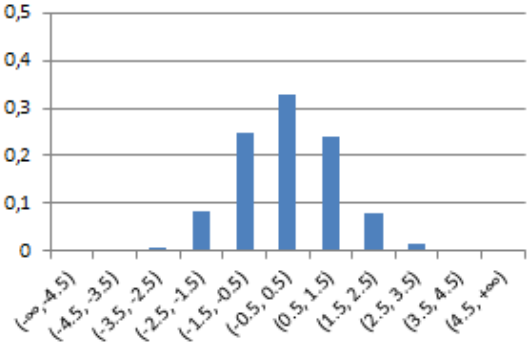
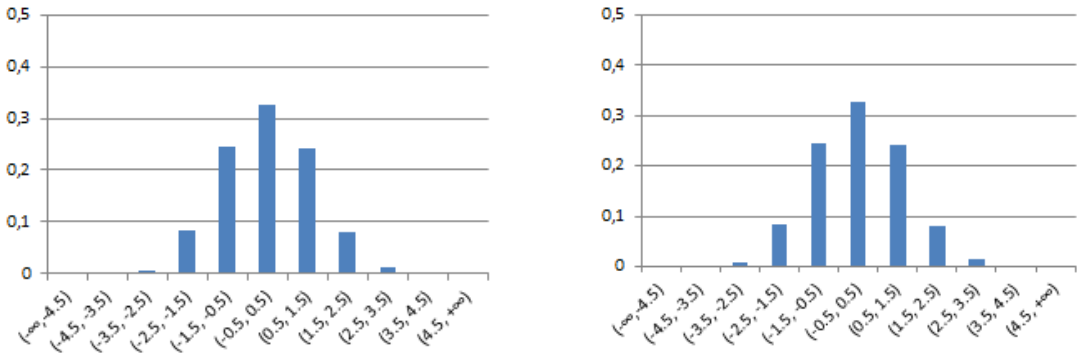


Figure 133. The prediction error histogram of the calibrated model over the measurements from the 4 different wafers used in the calibration process.

In order to validate the calibrated model, it was tested over two other wafers (presenting densities of 20% and 40% in the dummy blocks). Notice that those densities were not present in the calibration procedure. Both wafer measurements were predicted with RMS of 1.07 and 1.21nm, respectively. The histograms of the prediction errors are presented in Figure 134.



(a) (b)

Figure 134. The prediction error histograms of the calibrated model over the measurements from the 2 different wafers that were not used in the calibration process. (a) for the 20% dummy block wafer and (b) for the 40% dummy block wafer.

Notice that the histograms from both validation runs are very similar to the resulting from the calibration (shown in Figure 133).

In this exercise the modeling of extreme-long-range phenomena related to e-beam direct write lithography was demonstrated. The presented data indicate an impact of over 15% in CD for 65nm half-pitch technology. It may be expected to be of around 30% for 28nm half pitch and of around 40% for 22nm half pitch. Finally, 1 Gaussian PSF model was calibrated and validated, presenting a RMS value of around 1.1nm, in comparison with the process variability of 0.82nm. By employing this model and the technique for correction [FIGUEIRO, 2013c], also developed during the period of this thesis work, one may achieve better exposure control both for wafer direct writing and for mask writing.

7.4. Use of Cross-validation

In order to demonstrate the importance of performing cross-validation in order to certify the quality of a calibrated model, two tests were performed. Each test consisted of performing the calibration of a 3 Gaussian PSF with a constant threshold over a different calibration layout. The layouts were exposed using a Vistec SB3054 tool and measured using a SEM CG4000 from Hitachi. First, a calibration using all set of data available was used. Later, a cross-validation was performed. The cross-validation runs the calibration over ¾ of the patterns available and validate on the last ¼. The procedure is repeated 4 times so every pattern was used once for validation.

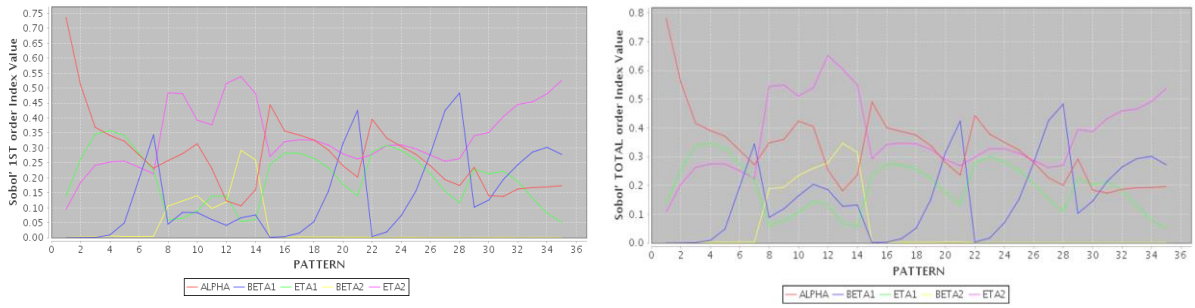
The first calibration procedure was performed based on the measurements over a layout containing 35 pattern types exposed at several different doses, totaling 329 different patterns. The patterns were LSB with CDs varying from 65nm to 250nm and the densities from 30% to 60%, ISO patterns and iISO patterns. The value of the parameters of the 3 Gaussian PSF model with constant threshold was: $\alpha = 37.56$nm, $\beta_1 = 438.86$nm, $\beta_2 = 8482.67$nm, $\eta_1 = 0.15$ and $\eta_2 = 0.12$. The resist threshold (TH) calibrated value was 70.17 µC/cm², in absolute value. The calibration cost function used was the RMSE, which resulted in a value of 3.02nm. The obtained RMSE is acceptable since it is of the same order of the variability observed on the data (around 3nm). The cross-validation result is shown in Table 34.

Table 34. Results of the four rounds of cross-validation for the first layout.

#	α (nm)	β_1 (nm)	β_2 (nm)	η_1 (nm)	η_2 (nm)	TH (μC/cm²)	RMSE (nm)
1	36.64	429.66	8210.89	0.17	0.11	69.90	3.03
2	38.95	353.66	10015.41	0.14	0.12	70.35	3.01
3	36.69	394.99	7794.68	0.16	0.12	70.01	2.98
4	36.53	305.24	8568.02	0.15	0.13	69.97	2.98

Observe that the result shows little variation on several parameters, such as the TH, η_1 and η_2. However, it is possible to observe that β_2 presents a very high variation. Since the RMSE is not varying much, it is a clear indicator that it is the calibration layout that lacks the ability to provide the information required to determine the value of this parameter.

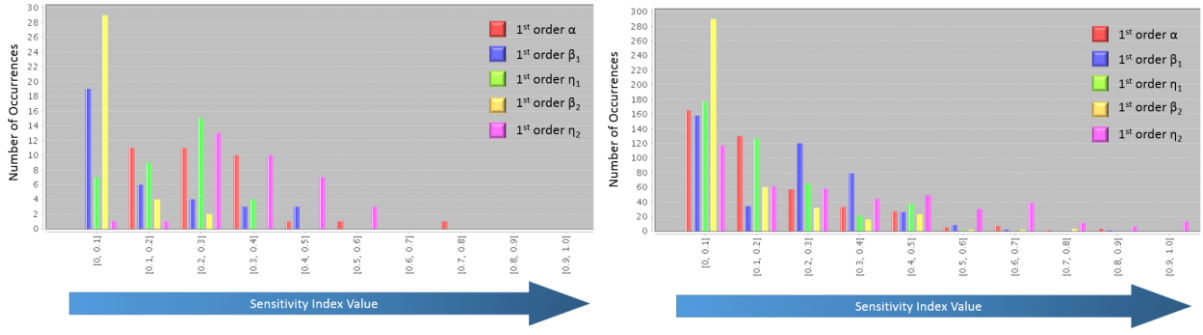
The pattern sensitivity was computed for comparison purposes, both 1st Order and Total Order Sobol' indices (as shown in Figure 135). Observe that β_2 presents only two patterns with 1st Order (Figure 135(a)) sensitivity index above 0.15. Moreover, even then Total Order (Figure 135(b)) is low for the β_2 parameter. Finally, other parameters present significantly higher sensitivity on the same pattern, which is not the ideal case for a calibration.



(a) (b)

Figure 135. Sobol' indices for sensitivity of the parameters of a 3 Gaussian PSF model over the patterns in the first layout (a) 1st Order and (b) Total Order.

Based on these results and in order to demonstrate that adding patterns with high sensitivity to all parameters improved calibration, a second calibration layout was proposed. This layout contained ISO and iISO patterns similar to the previous one, plus more different densities of LSB and several different variations of LBP patterns (which were absent from the first calibration pattern). The total number of patterns was 642. In comparison with the first calibration layout, one may see the sensitivity indices as shown in Figure 136.



(a) (b)

Figure 136. First order Sobol' indices for sensitivity of the parameters of a 3 Gaussian PSF model over the patterns in of (a) the first calibration layout and (b) the second calibration layout.

From the sensitivity indices from both calibration patterns presented in Figure 136 one may observe that the second one presents at least 20 patterns with sensitivities between 0.4 and 0.5 and at least one pattern whose 1st order Sobol' index is above 0.5. In comparison, the first layout had no pattern with sensitivity above 0.3 for β_2 and no pattern with sensitivity above 0.4 for η_1.

Using the metrology result after exposure of the second calibration layout, a 3 Gaussian PSF and a constant threshold was calibrated. Unfortunately, the exposure and measurement of this layout occurred in different conditions, which makes direct comparison of the parameters found impossible. The parameters obtained were the following: α = 25.89nm, β_1 = 121.12nm, β_2 = 9945.35nm, η_1 = 0.19 and η_2 = 0.37. The resist threshold used was a constant one, calibrated to the value of 63.42. The calibration cost function used was the RMSE, which provided a value of 3.48nm. The cross-validation result is shown in Table 35.

Table 35. Results of the four rounds of cross-validation for the second layout.

#	α (nm)	β_1 (nm)	β_2 (nm)	η_1 (nm)	η_2 (nm)	TH (µC/cm²)	RMSE (nm)
1	26.03	172.23	11237.45	0.19	0.37	62.59	3.32
2	27.20	139.68	10286.24	0.11	0.35	64.06	3.61
3	27.52	151.47	10418.51	0.15	0.38	63.73	3.47
4	26.75	168.52	10751.46	0.17	0.36	63.16	3.48

The results show little variation on most parameters, but a significant change on η_1, where the parameter may go from 0.19 to 0.11, according to the calibration run. The probable reason for this is the correlation from this parameter and β_1, which seems to be directly compensating this variation. Those are the only two parameters presenting significant value variation (above 10%) from one round to the next of the cross validation. The correlation among η_1 and β_1 is shown in the plot shown in Figure 137. The resulting R^2 is of 0.9458.

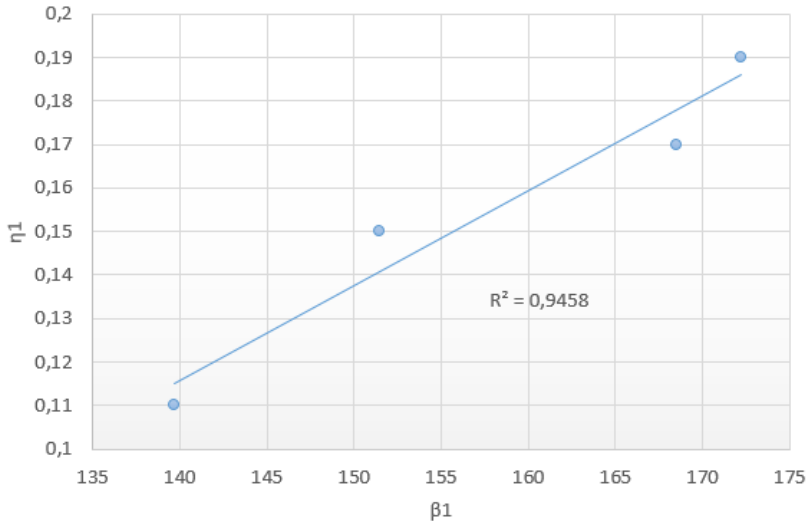


Figure 137. Evaluation of the correlation among the parameters η_1 and β_1 in the four rounds of cross-validation. R^2 is of 0.9458.

It is interesting to notice that performing cross-validation procedures on the two results obtained after these calibration exercises allowed to detect possible weaknesses on the generated models. Moreover, these results stress the importance of employing adequate calibration layout in order to obtain robust calibration results.

7.5. Sensitivity Analysis Characteristics

In order to present some specificities of the sensitivity analysis applied to calibration layouts for e-beam process modeling, some tests were performed and are presented in this section.

7.5.1. Influence of dominating parameters over Sensitivity Analysis indices

According to the structure of the model used and the tests performed, one may realize that one or more parameters dominate the influence over the variation of the output of the model. This is the case of the resist threshold when compared to the parameters of a PSF. When performing variance-based sensitivity analysis, such as the Sobol' indices used in this thesis, what one is evaluating is the relative impact of each parameter over the value of the output. One may be tempted to use the sensitivity analysis to evaluate all the aforementioned parameters at once. In this example we try to demonstrate the negative impact of doing so.

The procedure was the following. The Sobol' Indices were calculated over a simple layout for two different models, one composed of a 3 Gaussian PSF model and a constant threshold while the second presenting the same PSF but a fixed threshold at 0.5. The ranges of variation for each parameter is shown in Table 36.

Table 36. Range of variation for each parameter of the 3 Gaussian PSF.

Parameter	Minimum Value	Maximum Value
α	25	35
β_1	300	600
η_1	0.1	0.5
β_2	6000	10000
η_2	0.2	0.6

Moreover, for the analysis considering the sensitivity to the threshold, the variation of the parameter was considered from 0.35 to 0.65. The histogram of the distribution of patterns based on their sensitivity values is shown in Figure 138 (a) for the analysis including the threshold and Figure 138 (b) excluding the resist threshold parameter.

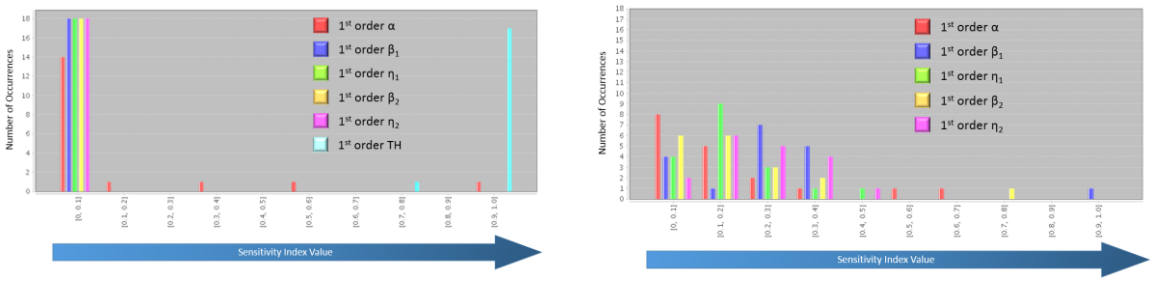


Figure 138. The estimated Sobol' indices histograms for 18 different patterns (a) including the analysis of the threshold value and (b) excluding the threshold value as a parameter under evaluation

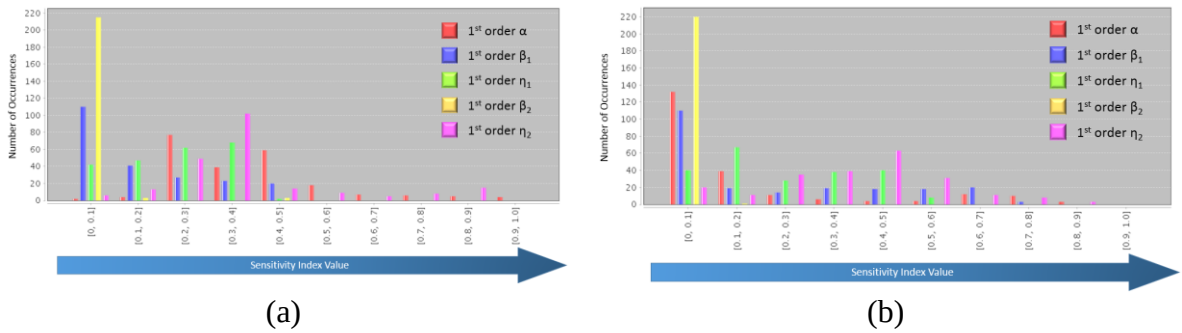
The results show that if the threshold parameter is taken into account, it dominates the sensitivity. It is an expected behavior since we may consider that the resist threshold presents impact over all patterns and any change on its value impact the estimated CD of all patterns. For this reason, one should not include the threshold in the evaluation of the sensitivity analysis of the parameters of the model.

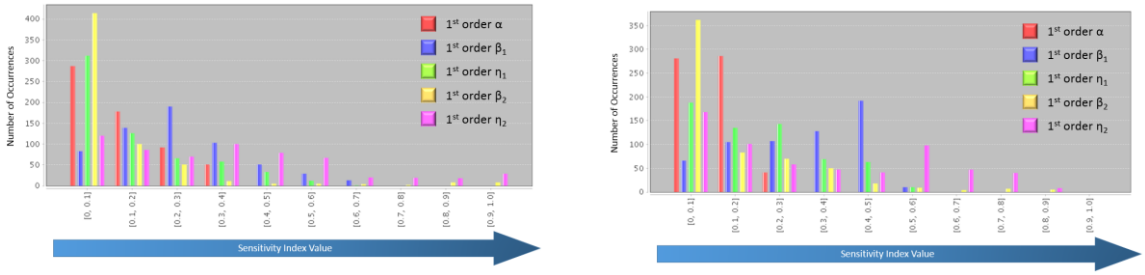
7.5.2. Impact of the pre-correction on the Sensitivity Analysis

Usually calibration layouts present patterns with very distinct characteristics in order to provide enough information for the calibration procedure. These patterns often present different CD values and different densities. Moreover, it is known that it is difficult to provide a model that explains the data in a very large range of variation and, for this reason, providing data not too far off from the target dimensions is advantageous for the calibration. This two conditions combined lead to the pre-correction of the calibration layout, which allows, for instance, patterns with high density to receive less dose than the ones with low density, enabling both to be not too far from the intended dimensions.

The purpose of this test is to verify the impact of performing pre-correction over the data over the SA indices. This is relevant since the SA indices are the main criteria while determining the patterns that should be included on the calibration layout.

The test consists in pre-correcting two test layouts, one composed on LSBs and another with LBPs and evaluate the sensitivity of the parameters of a 3 Gaussian PSF model. The results are shown in Figure 139.





(c) (d)

Figure 139. The estimated Sobol' indices histograms for (a) non-corrected LSB layout, (b) corrected LSB layout, (c) non-corrected LBP layout and (d) corrected LBP layout.

Although globally one may identify similar results for non-corrected and corrected patterns, some points are significantly different. First, when comparing the results from non-corrected and corrected LSB layouts (Figure 139 (a) and (b), respectively), one may see that the correction reduced significantly the sensitivity to the α parameter for several patterns. Moreover, there is a larger set of patterns sensitive to η_2 after the correction. In a similar fashion, the results from non-corrected and corrected LBP layouts (Figure 139 (c) and (d), respectively) show a significant difference. The parameter β_1 benefits of a significant improvement in sensitivity after correction while the α parameter is negatively impacted.

Since the sensitivity indices may vary for the patterns with or without pre-correction, one should always evaluate the sensitivity of the patterns as they are going to be exposed. If a pre-correction is going to be necessary to assure the quality of the exposure, the selection should be performed after the pre-correction. This is also true for the cases where a tool PEC algorithm is ON. Ideally, the algorithm should be emulated before running SA over the patterns.

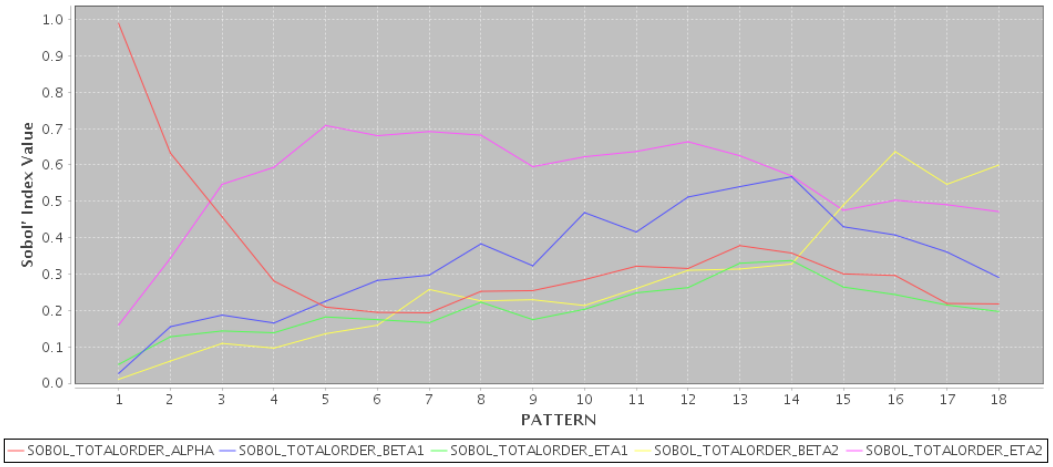
7.5.3. Sensitivity Analysis and the Range of the Parameters

Another relevant aspect while performing sensitivity analysis is the range of values considered to the parameters. One may think that using a large range that ensures the parameter to be covered is enough, but this is only the case if the values are realistic. To illustrate this issue, consider the computation of the sensitivity analysis indices of a 3 Gaussian PSF over 18 LSB patterns with CDs varying from 30nm to 200nm. Two runs are performed with the only difference of the change on the range of the αparameter. The parameter ranges of both runs is presented in Table 37.

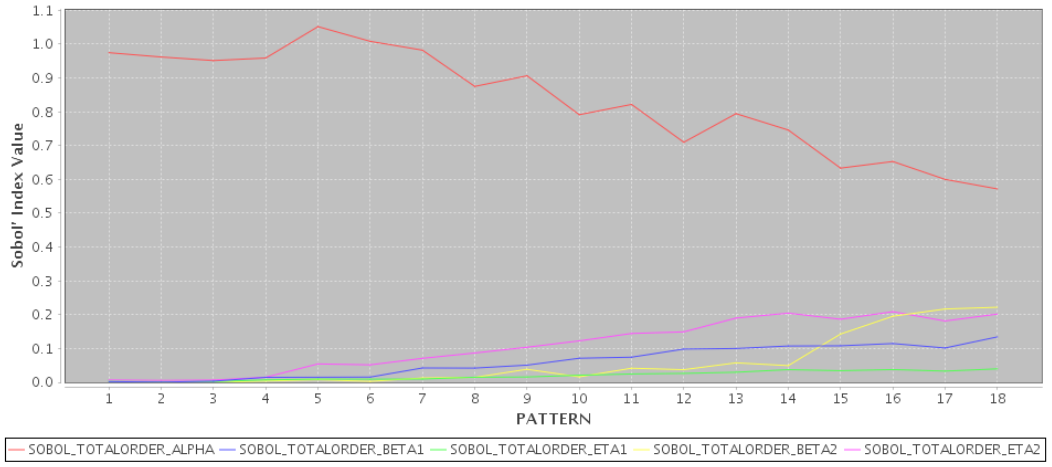
Table 37. Range of variation for each parameter for both runs.

Parameter	Run 1	Run 2
α	[25, 35]	[10, 100]
β1	[300, 600]	
η1	[0.1, 0.5]	
β2	[6000, 10000]	
η2	[0.2, 0.6]	

The sensitivity indices are presented in Figure 140 (a) for the first run, with the regular ranges and in Figure 140 (b) for the run with an enlarged range for the α parameter.



(a)



(b)

Figure 140. The estimated Sobol' indices values for 18 LSB patterns (a) with a regular range for all parameters and (b) with an large range for the α parameter.

The resulting sensitivity indices show that not only the sensitivity to α is impacted by the miss selection of the range, it also masks the information about the other parameters. For this reason, identifying the range of values that each parameter may assume is a key aspect of the sensitivity analysis and should be performed carefully to prevent results such as the ones presented on this section.

Conclusion

"After climbing a great hill, one only finds that there are many more hills to climb."

- *Nelson Mandela*

Conclusion

The study presented in this thesis addresses the relevant aspects of modeling e-beam lithography processes. Such models may be used both to improve resolution and to reduce writing time, for instance.

State-of-the-art strategies for todays and future lithography processes were presented and each of their main steps were described. First Principle models that explain the physics and chemistry of the most influential steps in the process resolution were also discussed. Moreover, general Compact models for predicting the results from e-beam lithography were also presented. Finally, some of the limitations of the current approach were described.

One result of the research developed during this thesis work was the proposition of new Compact models to achieve higher precision in the description of the e-beam lithography process. First, new point-spread functions (PSF) for electron scattering were proposed. A combination of Gaussian and Gamma distributions was calibrated based on the observations of backscattering phenomenon relying on First Principle models from Monte Carlo simulation. The results showed that the Gamma PSF was able to provide a better fit (lower error) than the standard Gaussian PSF with the same number of parameters.

A Voigt/Pearson type VII distribution was also proposed to explain electron scattering. A Pearson type VII distributions provided extra flexibility to the traditional Gaussian PSF. Finally, a third electron scattering model was proposed based on the flexibility of the Splines. All these PSF models are now proprietary to Aselta Nanographics and are implemented in Aselta Inscale™.

The basic concepts involving modeling were presented in this thesis. Topics such as data quality, model selection and model validation were introduced as tools to be applied in the process of modeling e-beam lithography. Moreover, the concepts of local and global sensitivity analysis were used extensively. Different strategies of global sensitivity analysis were presented and discussed as well as one of the main aspects in its evaluation, which is the space sampling approach. Finally, it was discussed why in the context of e-beam lithography the approach of choice is the variance-based sensitivity analysis and some of the strategies to estimate it were presented.

Moreover, a variable threshold resist model was proposed based on a surface described by metrics computed as convolution kernels. By adopting these metrics it was possible to propose an improved resist model that may also lump etch effects when only after etch measurements are available. Results using such method further improved the Pearson distribution approach on the critical features. The developed technology for computing a fast and robust kernel and the surface model interpolated based on this information are also proprietary to Aselta Nanographics.

The calibration of the proposed models was validated in an industrial production environment. The generated model used a Pearson PSF and two different strategies of threshold model. The first approach used a constant threshold and presented a standard deviation of the errors on the order of the process variation observed. The variable threshold approach reduced even further the error, obtaining a standard deviation of 2.29nm. Moreover, this strategy was able to reduce the errors range for critical CDs. Results after correction using a constant threshold model showed an error range was of around -5.5nm to 8nm while the one using a variable threshold was of around -2nm to 3.5nm.

The fogging effect is also one of the sources of error during e-beam lithography. In this thesis the same was also used to select the patterns to the calibration of the fogging model. A full description of a calibration design was presented and the results obtained by employing such model and a strategy on how to compensate for fogging were also described. This strategy could be extended to other extreme long range effects. The presented strategy can be performed off-line, as oppose of the state-of-the-art internal machine correction. The advantage of employing off-line compensation relies on allowing a compensation using dose and/or geometry modulation.

In this thesis it was also presented the different strategies to determine the values of the parameters of an e-beam lithography model. The approaches consisted of either relying on First Principle models or in experimental data. The second is preferable since the performance of First Principle models are usually restrictive to employing it in large volume of data and the lack of all the required information to obtain precise simulations. Moreover, by relying on experimental data, all the phenomena (even if not fully understood or predicted) is taken into consideration.

However, the drawback of relying on experiments while performing model calibration is the limitation in the metrology accuracy. A study regarding the impact of process and metrology variability over the calibration result was presented, indicating the limits one may expect from the generated model according to the quality of the data used. Finally, techniques for assuring the quality of a model such as cross-validation were also presented and demonstrated in some real-life situations.

Part of the quality of the model and of the data used to calibrate it comes from the experiment performed, whenever the calibration relies on it. The calibration layout is a key factor for providing the optimization algorithm with the resources necessary to determine the values of each of the parameters of the model. Several strategies from the literature were briefly described before introducing one of the main propositions of this thesis, which is employing variance-based global sensitivity analysis to determine which patterns are more suitable to be used in the calibration of a model. It was demonstrated the impact of the lack of sensitivity on the patterns used for calibration over the final parameter values. A complete flow for selecting patterns for a calibration layout was presented and finally, some results of poor designing of the sensitivity analysis experiment were demonstrated in the results section.

During the development of the work of this thesis some topics were left aside since it would not be feasible in such short period to be addressed as their complexity requires. One of this topics is the evaluation of different optimization algorithms and their influence over the final result of the model. It is possible to imagine using sensitivity analysis and other mathematical tools to evaluate the effort a priori and either select the optimization algorithm or adapt its parameters. Another axis is to employ the sensitivity analysis result to perform a model calibration by parts, reducing the complexity and, by doing so, enabling the optimizer to provide a better solution in a shorter period of time. Yet other topics not discussed in this work were the impact of the cost function used over the calibration result and adapted strategies for detecting outliers in metrology results.

Prediction is very difficult, especially about the future, would have said Niels Bohr. The art of modeling has so many factors that may lead the entire exercise to failure that is almost hard to explain why it often works. Nonetheless, some general remarks may be highlighted from the work of modeling e-beam lithography processes. The data provided for the calibration procedure has a determinant role in the quality of the generated model. Designing the experiment carefully according to the requirements and characteristics of the target model and the limitations of the process used is a key factor for a well succeeded calibration. Moreover, providing models that are able to explain the phenomena and using the tools for comparing models and preventing overfitting, for instance, are concerns that must be on the top of the list of every modeler. Finally, once the model is created, it is necessary to assure its quality by means of validation, first relying on the data already available and later by performing new experiments. This way, this hard task that is predicting the future becomes little less difficult.

Glossary

Alpha (α): Used in several point spread function models to represent the short range value of the model.

BEOL: Back-end-of-line

Beta (β): Used in several point spread function models to represent the long range value of the model.

CAR: Chemically amplified resist. A type of resist that, upon post-exposure bake, will multiply the number of chemical reactions by using chemical catalysis.

CD: Critical dimension, it corresponds to the smaller dimension of a given pattern.

CMP: Chemical-mechanical planarization

Corner Rounding: The rounding on the actual patterned feature of a nominally sharp, square corner on the layout.

CVD: Chemical vapor deposition

DoF: Depth-of-focus. The total range of focus that can be tolerated in terms of several different specifications.

DSA: Directed self-assembly

DUV: Deep ultra-violet. A term used to describe light of a wavelength in the range of around 150nm to 300nm.

DW: Direct write.

E-beam: Electron beam.

EOL: End-of-line. May stand for the measurement of shortening of the end of a line or for a calibration pattern composed of two blocks of lines whose ends are close to each other and their distance is measured.

Eta (η): Used in several point spread function models to represent the weight of long range in reference to the short range of the model.

EUV: Extreme ultra-violet

FEOL: Front-end-of-line.

High-Z material: Material composed of chemical element with a high atomic number.

IC: Integrated circuit.

iISO: Inverted isolated (line): It stands for a calibration pattern composed of an isolated trench between two pads. The space between the pads (the trench) is measured.

IQR: Interquartile range: The distance between the first and third quartiles of a distribution including 50% of the data centered around its mean value (second quartile).

ISO: Isolated (Line). It stands for a calibration pattern composed of an isolated line whose CD is measured.

ITRS: International Technology Roadmap for Semiconductors

LBP: Line between pads. A calibration pattern composed of a center line that is measured and two pads, one at each side of the line.

LER: Line-edge-roughness.

Low-Z material: material composed of chemical element with a low atomic number.

LSB: Line/Space Block. A grating of lines and spaces used as a calibration pattern.

MAD: Median Absolute Deviation. It is a measure of the variability of distribution around its median value.

Maskless lithography: Any strategy of lithography that does not require the use of a mask.

MSE: Mean square error.

MSLE: Mean square logarithmic error.

Mode: The value that appears more often in a series of values. In a distribution function, it represents the tallest peak.

Negative resist: A resist whose chemical structure allows for the unexposed areas to develop at a faster rate than the exposed areas.

NIL: Nano-imprint lithography

PAG: Photo acid generator.

PDF: Probability distribution function.

PEB: Post-exposure bake.

PEC: Proximity Effect Correction.

PVD: Physical vapor deposition.

Pitch: The sum of line width and space width for a repeating pattern.

Positive resist: A resist whose chemical structure allows for the exposed areas to develop at a faster rate than the unexposed areas.

PSF: Point Spread Function.

SEM: Scanning Electron Microscope. A machine based on using an electron-beam to bombard the material and perform measurement based on the images generated by the backscattered electrons.

SOI: Silicon-on-insulator.

Target CD: Intended dimension of a given pattern on the layout.

Target Layout: Typical layout intended to be corrected using an e-beam model. Usually a real application such as memory chips or random logic.

Technology Node: Technology node or just node is a reference to the smallest CD a process may print. In some other texts one may find its definition to be the half-pitch distance or the transistor gate size of a process.

Yield: The fraction of die that work properly at the end of fabrication.

WPH: Wafers per hour. Usually used in the context of measuring the speed of a machine or of the entire flow.

Publications

Papers

"Improved electron backscattering representation using a new class of distribution: application to EUV masks"
T. Figueiro, M. Saib, N. Jedidi, P. Schiavone
Proceedings of SPIE 8323, 2012.

"Advanced module for model parameter extraction using global optimization and sensitivity analysis for electron beam proximity effect correction"
T. Figueiro, K-H Choi, M. Gutsch, M. Freitag, C. Hohle, J-H Tortai, M. Saib, P. Schiavone
Proceedings of SPIE 8522, 2012.

"Sensitivity analysis for accurate determination of PSF parameters"
T. Figueiro, J-H Tortai, P. Schiavone
Microelectronic Engineering Volume 97, 2012.

"PSF calibration patterns selection based on Sensitivity Analysis"
T. Figueiro, M. Saib, J-H Tortai, P. Schiavone
Microelectronic Engineering Volume 112, 2013.

"Reticle level compensation for long range effects"
T. Figueiro, C. Browning, M. J. Thornton, C. Vannuffel, P. Schiavone
Proceedings of SPIE 8680, Alternative Lithographic Technologies V, 2013.

"Extreme long range process effects characterization and compensation"
T. Figueiro, C. Browning, M. J. Thornton, C. Vannuffel, P. Schiavone
Proceedings of SPIE 8886, 29th European Mask and Lithography Conference, 2013.

"Metrology variability and its impact in process modeling"
T. Figueiro, M. Saib, K-H Choi, C. Hohle, M. J. Thornton, C. Vannuffel, J-H Tortai, P. Schiavone
Proceedings of SPIE 8880, Photomask Technology, 2013.

"Effects on electron scattering and resist characteristics using assisting underlayer for e-beam direct write lithography"
X. Thrun, K-H. Choi, N. Hanisch, C. Hohle, J. Steidel, D. Guerrero, T. Figueiro, J. W. Bartha
Proceedings of SPIE 8682, 2013.

"Demonstration of EDA flow for massively parallel e-beam lithography"
P. Brandt, J. Belledent, C. Tranquillin, T. Figueiro, S. Meunier, S. Bayle, A. Fay, M. Milléquant, B. Icard, M. Wieland
Proceedings of SPIE 9049, 2014.

"Effective corner rounding correction in the data preparation for electron beam lithography"
K-H. Choi, C. Browning, T. Figueiro, C. Hohle, M. Kaiser, P. Schiavone
Proceedings of SPIE. 9235, Photomask Technology, 2014.

"Photonic curvilinear data processing"

C. Browning, T. Quaglio, T. Figueiro, S.P Pauliac, J. Belledent, A. Fay, J. Bustos, J-C. Marusic, P. Schiavone
Proceedings of SPIE. 9235, Photomask Technology, 2014

Patents

“Procédé de correction des effets de proximité électronique utilisant des fonctions de diffusion décentrées”,
P. Schiavone, T. Figueiro
Patent filed, 1157338, 2011. (in French)

“Procédé de correction des effets de proximité électronique utilisant des fonctions de diffusion de type voigt”
J-H. Tortai, P. Schiavone, T. Figueiro, N. Jedidi
Patent filed, 1253389, 2012. (in French)

“Procédé de correction des effets de proximité électronique”
N. Jedidi, P. Schiavone, J-H. Tortai, T. Figueiro
Patent filed, 1302001, 2013. (in French)

“Method for Determining the Parameters of an IC Manufacturing Process by a Differential Procedure”
M. Saib, P. Schiavone, T. Figueiro
Patent filed, 14305835.2, 2014.

“Method for Determining the Parameters of an IC Manufacturing Process”
M. Saib, P. Schiavone, T. Figueiro
Patent filed, 14306241.2, 2014.

“Method for Calculating the Metrics of an IC Manufacturing Process”
M. Saib, A. Fay, P. Schiavone, T. Figueiro
Patent filed, 14305834.5, 2014.

References

[ABE, 2007] T. Abe, J. I. Suzuki, J. Yashima, T. Iijima, S. Oogi, H. Anze, Y. Onimaru, H. Tsurumaki, S. Tsuchiya, Y. Hattori, "Global Critical Dimension Correction: I. Fogging Effect Corretion", Japanese Journal of Applied Physics, Vol. 46, No. 6A, 3359-3367, 2007.

[AKAIKE, 1973] H. Akaike, Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki (Eds.), Second international symposium on information theory, 267-281. Budapest: Academia Kiado, 1973.

[ARTHUR, 1997] G. Arthur, N. Eilback, B. Martin, "Effect of temperature variations in the post exposure processes of optical lithography", Microelectronic Engineering, 35, 137-140, 1997.

[ARTHUR, 1998] G. Arthur, C. A. Mack, N. Eilback, B. Martin, "Analyzing the dissolution characteristics of deep UV chemically amplified photoresist", Microelectronic Engineering, 41/42, 311-314, 1998.

[AYDIN, 2013] D. Aydin, M. Memmedli, R. E. Omay, "Smooting Parameter Selection for Nonparametric Regression Using Smoothing Spline", European Journal of Pure and Applied Mathematics, Vol. 6, No. 2, 222-238, 2013.

[BABIN, 1996] S. Babin, "Method for Measuring Absorbed Energy Density in Photoresist in a Laser Pattern Generation", Proceeding of SPIE 2726, 859-865, 1996.

[BABIN, 2003] S. V. Babin, S. Borisov, E. Cheremukhin, E. Grachev, V. Korol, L. E. Ocola, "Software tool for advanced Monte Carlo simulation of electron scattering in EBL and SEM: CHARIOT", Proceedings of SPIE 5037, 2003.

[BAKSHI, 2009] V. Bakshi, "EUV Lithography", John Wiley & Sons, 2009.

[BRANDT, 2014] P. Brandt, J. Belledent, C. Tranquillin, T. Figueiro, S. Meunier, S. Bayle, A. Fay, M. Milléquant, B. Icard, M. Wieland, "Demonstration of EDA flow for massively parallel e-beam lithography", Proceedings of SPIE 9049, 2014.

[BURDORF, 2000] J. Burdorf, C. Pierrat, "Method for Multiple Process Parameter Matching", Patent No. US 6,033,814, 2000.

[BURDORF, 2002] J. Burdorf, C. Pierrat, "System and Method for Process Matching", Patent No. US 6,463,403 B1, 2002.

[BURNHAM, 2002] K. P. Burnham, D. R. Anderson, "Multimodel Inference – Understanding AIC and BIC in Model Selection", Second Edition, Springer, 2002.

[BURNHAM, 2004] K. P. Burnham, D. R. Anderson, “Model Selection and Multimodel Inference – A Practical Information-Theoretic Approach”, Sociological Methods & Research, Vol. 33, No. 2, 261-304, 2004.

[CACUCI, 2003] D. G. Cacuci, “Sensitivity and Uncertainty Analysis, vol. 1”, Chapman and Hall Publisher, 2003.

[CAMPBELL, 2006] K. Campbell, M. D. McKay, and B. J. Williams, “Sensitivity analysis when model outputs are functions”, Reliability Engineering and System Safety, 91, 1468–1472, 2006

[CAMPOLONGO, 2007] F. Campolongo, J. Cariboni, A. Saltelli, “An effective screening design for sensitivity analysis of large models”, Environmental Modelling and Software 22, 1509-1518, 2007.

[CHANG, 1975] T. H. P. Chang, “Proximity effect in electron-beam lithography”, J. Vac. Sci. Technol., Vol. 12, No. 6, 1271-1275, 1975.

[CHOU, 1996] S. Y. Chou, P. R. Krauss, P. J. Renstrom, “Imprint Lithography with 25-Nanometer Resolution”, Science, New Series, Vol. 272, Issue 5258, 85-87, 1996.

[COBB, 1997] N. Cobb, A. Zakhor, M. Reihani, F. Jahansooz, V. Raghavan, “Experimental Results on Optical Proximity Correction with Variable Threshold Resist Model”, Proceedings of SPIE 3051, 458-468, 1997.

[COBB, 1998] N. Cobb, “Fast Optical and Process Proximity Correction Algorithms for Integrated Circuit Manufacturing”, University of California at Berkeley, PhD Thesis, 1998.

[CRAIGHEAD, 1983] H. G. Craighead, R. E. Howard, L. D. Jacket and P. M. Mankiewich, “10-nm linewidth electron beam lithography on GaAs”, Applied Physics Letters 42; 38-40, 1983.

[CROFFIE, 2000] E. Croffie, M. Cheng, A. Neureuther, R. Cirelli, F. Houliban, J. Sweeney, P. Watson, O. Nalamasu, I. Rushkni, O. Dimov, A. Gabor, “Overview of the STORM Program Application to 193nm Singe Layer Resists”, Microelectronic Engineering, 53, 437-442, 2000.

[DAMBLIN, 2013] G. Damblin, M. Couplet, B. Iooss, “Numerical studies of space filling designs: optimization of Latin hypercube samples and subprojection properties”, Journal of Simulation, 7, 276-289, 2013.

[DENNARD, 1974] R. H. Dennard, F. H. Gaensslen, H. Yu, V. L. Rideout, E. Bassous and A. R. LeBlanc, “Design of Ion-Implanted MOSFET’s with Very Small Physical Dimensions”, IEEE Journal of Solid-state Circuits, Vol. SC-9, No. 5, 256-268, 1974.

[DEVROYE, 1979], L. P. Decroye, T. J. Wagner, “Distribution-free performance bounds for potential function rules”, IEEE Transactions Information Theory, Vol. 25, No. 5, 601-604, 1979.

[DIAL, 1998] O. Dial, C. C. Cheng and A. Scherer, “Fabrication of high-density nanostructures by electron beam lithography”, J. Vac. Sci. Technology, B 16, 3887-3890, 1998.

[DILL, 1975] F. H. Dill, W. P. Hornberger, P. S. Hauge, J. M. Shaw, “Characterization of Positive Photoresist”, IEEE Transactions on Electron Devices, Vol. ED-22, No. 7, 445-452, 1975.

[DROUIN, 1997] D. Drouin, P. Hovington, R. Gauvin, “CASINO: A new Monte Carlo code in C language for electron beam interactions – part II: Tabulated Values of the Mott Cross Section”. Scanning, Vol. 19, Issue 1, 20-28, 1997.

[DROUIN, 2007] D. Drouin, A. R. Couture, D. Joly, X. Tastet, V. Aimez, R. Gauvin, “CASINO V2.42 – A Fast and Easy-to-use Modeling Tool for Scanning Electron Microscopy and Microanalysis Users”. Scanning, Vol. 29, 92-101, 2007.

[ERDMANN, 2001] A. Erdmann, W. Henke, S. Robertson, E. Richter, B. Tollkühn, W. Hoppe, “Comparison of Simulation Approaches for Chemically Amplified Resists”, Proc. Of SPIE Vol. 4404, 99-110, 2001.

[FADEEVVA, 1954] V. N. Fadeevva, and N. M. Terent’ev, “Tables of Values of the Probability Integral for Complex Arguments”, State Publishing House for Technical Theoretical Literature, Moscow.

[FANG, 2006] K-T. Fang, R. Li and A. Sudjianto, “Design and modeling for computer experiments”. Chapman & Hall / CRC, 2006.

[FAURE, 2012] T. Faure, B. Iooss, H. Monod, L. Rouan, “Echantillonnage en grande dimension pour la propagation d’incertitude” in “*Analyse de sensibilité et exploration de modèles. Applications aux modèles environnementaux*”, R. Faivre, B. Iooss, S. Mahévas, D. Makowski, H. Monod, editors. Quae éditions, 2012. (*in French)*

[FEUILLARD, 2007] V. Feuillard. “Analyse d’une base de données pour la calibration d’un code de calcul”. Université Pierre et Marie Curie, Paris 6, PhD Thesis, 2007. (*in French)*

[FIGUEIRO, 2012a] T. Figueiro, M. Saib, N. Jedidi, P. Schiavone, “Improved electron backscattering representation using a new class of distribution: application to EUV masks”, Proc. SPIE 8323, Alternative Lithographic Technologies IV, (2012)

[FIGUEIRO, 2012b] T. Figueiro, K-H Choi, M. Gutsch, M. Freitag, C. Hohle, J-H Tortai, M. Saib, P. Schiavone, “Advanced module for model parameter extraction using global optimization and sensitivity analysis for electron beam proximity effect correction”, Proc. SPIE 8522, Photomask Technology 2012, (2012)

[FIGUEIRO, 2012c] T. Figueiro, J-H Tortai, P. Schiavone, “Sensitivity analysis for accurate determination of PSF parameters”, Microelectronic Engineering Volume 97, September 2012, (2012)

[FIGUEIRO, 2013a] T. Figueiro, M. Saib, J-H Tortai, P. Schiavone, “PSF calibration patterns selection based on Sensitivity Analysis”, Microelectronic Engineering Volume 112, December 2013, (2013)

[FIGUEIRO, 2013b] T. Figueiro, C. Browning, M. J. Thornton, C. Vannuffel, K-H Choi, C. Hohle, J-H Tortai, P. Schiavone, “Reticle level compensation for long range effects”, Proc. SPIE 8680, Alternative Lithographic Technologies V, (2013)

[FIGUEIRO, 2013c] T. Figueiro, C. Browning, M. J. Thornton, C. Vannuffel, P. Schiavone, “Extreme long range process effects characterization and compensation”, Proc. SPIE 8886, 29th European Mask and Lithography Conference, (2013)

[FIGUEIRO, 2013d] T. Figueiro, M. Saib, K-H Choi, C. Hohle, M. J. Thornton, C. Vannuffel, J-H Tortai, P. Schiavone, “Metrology variability and its impact in process modeling”, Proc. SPIE 8880, Photomask Technology 2013, (2013).

[FONSECA, 2009] C. Fonseca, M. Somervell, S. Scheer, W. Printz, K. Nafus, S. Hatakeyama, Y. Kuwahara, T. Niwa, S. Bernard, R.Gronheid, “Advances and challenges in dual-tone development process optimization”, Proceedings of SPIE 7274, Optical Microlithography XXII, 2009.

[GAMBOA, 2013] F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. “Sensitivity indices for multivariate outputs”. Comptes Rendus de l’Académie des Sciences, 351, 307–310, 2013.

[GEISSER, 1993] S. Geisser, “Predictive Inference”, Chapman and Hall, 1993.

[GRANIK, 2002] Y. Granik, N. Cobb, T. Do, “Universal Process Modeling with VTRE for OPC”, Proceedings of SPIE 4691, 377-394, 2002.

[GRANIK, 2002b] Y. Granik, N. Cobb, “MEEF as a Matrix”, Proceedings of SPIE 4562, 2002.

[GRANIK, 2003] Y. Granik, “Dry etch proximity modeling in mask fabrication”, Proceedings of SPIE 5130, 86-91, 2003.

[GRUBBS, 1969] F. E. Grubbs, “Procedures for Detecting Outlying Observations in Samples”, Technometrics, Vol. 11, No. 1, 1-21, 1969.

[HAUPTMANN, 2009] M. Hauptmann, K-H. Choi, P. Jaschinsky, C. Hohle, J. Kretz, L. M. Eng, “Determination of proximity effect parameters by means of CD-linearity in sub 100 nm electron beam lithography”, Microelectronic Engineering 86, 539-543, 2009.

[HENKE, 1999] W. Henke, M. Torkler, “Modeling of edge roughness in ion projection lithography”, J. Vac. Sci. Technol. 17, 3112-3118, 1999.

[HICKERNELL, 1998] F. J. Hickernell. “A generalized discrepancy and quadrature error bound”. Mathematics of Computation, 67, 299-322, 1998.

[HOVINGTON, 1997a] P. Hovington, D. Drouin, R. Gauvin, “CASINO: A new Monte Carlo code in C language for electron beam interactions – part I: Description of the program”. Scanning, Vol. 19, Issue 1, 1-14, 1997.

[HOVINGTON, 1997b] P. Hovington, D. Drouin, R. Gauvin, D. C. Joy, N. Evans, “CASINO: A new Monte Carlo code in C language for electron beam interactions – part III: Stopping power at low energies”. Scanning, Vol. 19, Issue 1, 29-35, 1997.

[HUDEK, 2006] P. Hudek, D. Beyer, “Exposure optimization in high-resolution e-beam lithography”, Microelectronic Engineering 83, 780-783, 2006.

[HUDEK, 2007] P. Hudek, U. Denker, D. Beyer, N. Belic, and H. Eisenmann, “Fogging effect correction method in high-resolution electron beam lithography”. Microelectron. Eng. 84, 5-8, 814-817, 2007.

[HURVICH, 1989] C. M. Hurvich, C-L. Tsai, “Regression and Time Series Model Selection in Small Samples”, Biometrika, Vol. 76, No. 2, 297-307, 1989.

[IOOSS, 2015] B. Iooss, P. Lemaître, “A review on global sensitivity analysis methods” In Uncertainty management in “Simulation-Optimization of Complex Systems: Algorithms and Applications”, C. Meloni and G. Dellino (eds), Springer, 2015.

[ITRS, 2013] International Technology Roadmap for Semiconductors, 2013 Edition, Lithography, 2013.

[KAHNG, 1999] A. Kahng, G. Robins, A. Singh, and A. Zelikovsky. “Filling Algorithms and Analyses for Layout Density Control”, IEEE Transactions on CAD, Vol. 18, No. 4, 445-462, 1999.

[KARAFYLLIDIS, 2000], I. Karafyllidis, P. I. Hagouel, A. Thanailakis, A. R. Neureuther, “An Efficient Photoresist Development Simulator Based on Cellular Automata with Experimental Verification”, IEEE Transactions on Semiconductor Manufacturing, Vol. 13, No. 1, 61-75, 2000.

[KURIHARA, 1995] K. Kurihara, K. Iwadate, H. Namatsu, M. Nagase, H. Takenaka, K. Murase, “An Electron Beam Nanolithography System and its Application to Si Nanofabrication”, Jpn. J. Appl. Phys. Vol 34, 6940-6946, 1995.

[JEDIDI, 2013] N. Jedidi, P. Schiavone, J-H. Tortai, T. Figueiro, “Procédé de correction des effets de proximité électronique”, Patent filed, 1302001, 2013. (in French)

[JEONG, 2013] S-J. Jeong, H. Y. Kim, B. H. Kim, H-S. Moon, S. O. Kim, “Directed self-assembly of block copolymers for next generation nanolithography”, Materials Today, Vol. 16, Issue 12, 468-476, 2013.

[JIN, 2005] R. Jin, W. Chen and A. Sudjianto. “An efficient algorithm for constructing optimal design of computer experiment”. Journal of Statistical Planning and Inference, 134, 268-287, 2005.

[JONES, 1998] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient Global Optimization of Expensive Black-Box Functions”, J. of Global Optimization, 13, 455-92, 1998.

[JOHNSON, 1990] M. E. Johnson, L.M. Moore and D. Ylvisaker. “Minimax and maximin distance designs”. Journal of Statistical Planning and Inference, 26, 131-148, 1990.

[KAMIKUBO, 2010] T. Kamikubo, T. Ohnishi, S. Hara, H. Anze, Y. Hattori, S. Tamamushi, S. Bai, J-S. Wang, R. Howell, G. Chen, J. Li, J. Tao, J. Wiley, T. Kurosawa, Y. Saito, T. Takigawa, “Mask Process Correction (MPC) modeling and its application to EUV mask for Electron beam mask writer, EBM-7000”, Proceedings of SPIE 7823, 2010.

[KASS, 1995] R. E. Kass, A. E. Raftery, “Bayes Factors”, Journal of the American Statistical Association, Vol. 90, No. 430, 773-795, 1995.

[KEIL, 2009] K. Keil, M. Hauptmann, K-H. Choi, J. Kretz, L. M. Eng, J. W. Bartha, “Fast backscattering parameter determination in e-beam lithography with a modified doughnut test”, Microelectronic Engineering 86, 2408-2411, 2009.

[KENDALL, 1938] M. G. Kendall, “A new measure of rank correlation”. Biometrika 30, 81-93, 1938.

[KIRKPATRICK, 1983] S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, “Optimization by Simulated Annealing”, Science 220 (4598), 671-680, 1983.

[KOHAVI, 1995] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”, International Joint Conference on Artificial Intelligence, 1995.

[KRASNOPEROVA, 1994] A. A. Krasnoperova, M. Khan, S. Rhyner, J. W. Taylor, Y. Zhu and F. Cerrina, “Modeling and simulations of a positive chemically amplified photoresist for x-ray lithography”, J. Vac. Sci. Technol. B 12(6), 3900-3904, 1994.

[KRATSCHMER, 1981] E. Kratschmer, “Verification of a proximity effect correction program in electron-beam lithography”, J. Vac. Sci. Technol. 19 (4), 1264-1268, 1981.

[KYSER, 1974] D. F. Kyser, K. Murata, “Quantitative Electron Microprobe Analysis of Thin Films on Substrates”, IBM Journal of Research and Development, Volume 18, Issue 4, 352-363, 1974.

[KYSER, 1975] D. F. Kyser, N. S. Viswanathan, “Monte Carlo simulation of spatially distributed beams in electron-beam lithography”, J. Vac. Sci. Technol., Vol. 12, No. 6, 1305-1308, 1975.

[KURIHARA, 1995] K. Kurihara, K. Iwadate, H. Namatsu, M. Nagase, H. Takenaka and K. Murase, “An Electron Beam Nanolithography System and its Application to Si Nanofabrication”, Jpn. J. Appl. Phys., 34, 6940–6946, 1995.

[LEE, 1981] K. L. Lee and H. Ahmed, “An e-beam microfabrication system for nanolithography”, J. Vac. Sci. Technology, 19, 946-949, 1981.

[LEVENBERG, 1944] K. Levenberg. “A method for the solution of certain non-linear problems in least squares”, Quarterly of Applied Mathematics 2,164–168, 1944.

[LIN, 2010] S. J. Lin, W. C. Wang, P. S. Chen, C. Y. Liu, T. N. Lo, J. H. Chen, F. Krecinic, B. J. Lin, “Characteristics performance of production-worthy multiple e-beam maskless lithography”, Proceedings of SPIE 7637, 2010.

[LIU, 2007], R. Liu, J. J. Zheng, H. J. Zhou, Y. C. Tian, G. Liu, L. G. Shen, “A Generalized Dill Exposure Model for Negative Thick Photoresist”, Proc. of IEEE, International conference on information acquisition, 285-291, 2007.

[MACK, 1995] C. A. Mack, “Lithography Effects of Acid Diffusion In Chemically Amplified Resists”, OCG Microlithography Seminar Interface ’95, 217-228, 1995.

[MACK, 1998] C. A. Mack, G. Arthur, “Notch Model for Photoresist Dissolution”, Electrochemical and Solid State Letters, Vol. 1, No. 2, 1998.

[MACK, 2007] C. A. Mack, “Fundamental Principles of Optical Lithography: The Science of Microfabrication”, John Wiley & Sons, 2007.

[MACK, 2011] C. A. Mack, “Fifty Years of Moore’s Law”, IEEE Transactions on Semiconductor Manufacturing, Vol. 24, No. 2, 202-207, 2011.

[MAILFERT, 2014] J. Mailfert, J. V. Kerkhove, P. Bisschop, K. Meyer, “Metal1 patterning study for random-logic applications with 193i, using calibrated OPC for Litho and Etch”, Proceedings of SPIE 9052, Optical Microlithography XXVII, 2014.

[MANAKLI, 2007] S. Manakli, C. Soonekindt, L. Pain, J-C. Le-Denmat, J. Todeschini, B. Icard, B. Minghetti, “Complementary dose and geometrical solutions for electron beam direct write lithography proximity effects correction: application for sub-45nm node product manufacturing”, J. Micro/Nanolith. MEMS MOEMS 6(3), 2007.

[MARA, 2008] T. A. Mara, S. Tarantola, “Application of Global Sensitivity Analysis of Model Output to Building Thermal Simulations”, Build Simul, 290-302, 2008.

[MARQUARDT, 1963] D. W. Marquardt, “An algorithm for the least-squares estimation of nonlinear parameters”, J. Appl. Math. 11, 431-441, 1963.

[MARREL, 2008] A. Marrel. “Mise en œuvre et utilisation du métamodèle processus gaussien pour l'analyse de sensibilité de modèles numériques : application à un code de

transport hydrogéologique". Institut National des Sciences Appliquées de Toulouse, PhD Thesis, 2008. *(in French)*

[MARREL, 2011] A. Marrel, B. Iooss, M. Jullien, B. Laurent, and E. Volkova. "Global sensitivity analysis for models with spatially dependent outputs". Environmetrics, 22, 383–397, 2011.

[MARTIN, 2011] L. Martin, "Méthodes de correction avancées des effets de proximité en lithographie électronique à écriture directe. Application aux technologies sub-32nm", Institut National des Sciences Appliquées de Lyon, PhD Thesis, 2011. *(in French)*

[MCKAY, 1979] M. D. Mckay, R. J. Beckman, W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code", Technometrics, Vol. 42, No. 1, 55-61, 1979.

[MILLER, 1998] J. H. Miller. "Active Nonlinear Tests (Ants) of Complex Simulation Models". Management Science, Volume 44, Issue 6, pages 551-555, 2000.

[MOORE, 1965] G. E. Moore, "Cramming more components onto integrated circuits", Electronics, Vol. 38, No. 8, 1965.

[MOORE, 1975] G. E. Moore, "Progress In Digital Integrated Electronics", Electron Devices Meeting, Vol. 21, 1975.

[MOORE, 1995] G. E. Moore, "Lithography and the Future of Moore's Law", Proceedings of SPIE, Vol. 2437, 1995.

[MOROKOFF, 1994] W. J. Morokoff, R. E. CAFLISCH, "Quasi-random sequences and their discrepancies", SIAM J. Sci. Comput. Vol. 15, No. 6, 1251-1279, 1994.

[MORRIS, 1991] M. D. Morris, "Factorial Sampling Plans for Preliminary Computational Experiments", Technometrics, Vol. 33, No. 2, 161-174, 1991.

[MÜLDERS, 2005] T. Mülders, W. Henke, K. Elian, C. Nölscher, M. Sebald, "New stochastic post-exposure bake simulation method", J. Microlith. Microfab., Microsyst., 4, 043010, 2005.

[MÜLLER, 2000] J. W. Müller. "Possible Advantages of a Robust Evaluation of Comparisons". Journal of Research of the National Institute of Standards and Technology, Volume 105, Number 4, pages 820-830, 1998.

[NAYLOR, 1967] T. H. Naylor, J. M. Finger, "Verification of Computer Simulation Models", Management Science, Vol. 2, 92-101, 1967.

[NELDER, 1965] J. A. Nelder and R. Mead, "A simplex method for function minimization", Computer Journal 7, 308-313, 1965.

[NIST, 2014] National Institute of Standards and Technology (NIST), "Engineering Statistics Handbook", http://www.itl.nist.gov/div898/handbook/index.htm, 2014.

[OWA, 2014] S. Owa, S. Wakamoto, M. Murayama, H. Yaegashi, K. Oyama, “Immersion lithography extension to sub-10nm nodes with multiple patterning”, Proceedings of SPIE 9052, Optical Microlithography XXVII, 2014.

[PEARSON, 1895] K. Pearson, “Notes on regression and inheritance in the case of two parents”. Proceedings of the Royal Society of London 58, 240-242, 1895.

[PETRIC, 2009a] P. Petric, C. Bevis, A. Carroll, H. Percy, M. Zywno, K. Standiford, A. Brodie, N. Bareket and L. Grella, “REBL: A novel approach to high speed maskless electron beam direct write lithography”, J. Vac. Sci. Technology, B 27, 161-166, 2009.

[PETRIC, 2009b] P. Petric, C. Bevis, A. Brodie, A. Carroll, A. Cheung, L. Grella , M. McCord, H. Percy, K. Standiford and M. Zywno, “REBL Nanowriter: Reflective Electron Beam Lithography”, Alternative Lithographic Technologies, Proc. Of SPIE Vol. 7271, 2009.

[PFEIFFER, 2010] H. C. Pfeiffer, “Direct write electron beam lithography: a historical overview”, Proc. Of SPIE Vol. 7823, 2010.

[PLATZGUMMER, 2013] E. Platzgummer, C. Klein, H. Loeschner, “Electron multi-beam technology for mask and wafer writing at 0.1nm address grid”, *J. Micro/Nanolith. MEMS MOEMS*. 12(3), 2013.

[POSTNIKOV, 2003] S. Postnikov, S. Hector, C. Garza, R. Peters, V. Ivin, “Critical dimension control in optical lithography”, Microelectrionic Engineering 69, 452-458, 2003.

[RANDALL, 1999] J. Randall, H. Gangala, A. Tritchkov, “Lithography Simulation with Aerial image – Variable Threshold Resist Model”, Microelectronic Engineering 46, 59-63, 1999.

[REINSCH, 1967] C. H. Reinsch, “Smooting by spline functions”, Numer. Math. Vol 10, no. 3, 177-183, 1967.

[RIO, 2010] D. Rio, “Lithographie électronique basse énergie: application au multifaisceaux”, Université de Grenoble – Institut Polytechnique de Grenoble, PhD Thesis, 2010. *(in French)*

[RISHTON, 1986] S. A. Rishton, D. P. Kern, “Point exposure distribution measurements for proximity correction in electron beam lithography on a sub-100 nm scale”, J. Vac. Sci. Technol. B 5 (1), 135-141, 1986.

[RISHTON, 1999] S. A. Rishton, J. K. Varner, L. H. Veneklasen, V. Boegly, A. L. Sagle, U. Hogmann, H. Kao, W. Wang, “Raster shaped beam pattern generation”, J. Vac. Sci. Technol. B17, No. 6, 2927-2931, 1999.

[SAHOURIA, 2000] E. Sahouria, Y. Granik, N. Cobb, O. Toublan, “Full-chip Process Simulation for Silicon DRC”, Modeling and Simulation of Microsystems, 32-35, 2000.

[SAIB, 2014a] M. Saib, P. Schiavone, T. Figueiro, “Method for Determining the Parameters of an IC Manufacturing Process by a Differential Procedure”, Patent filed, 14305835.2, 2014.

[SAIB, 2014b] M. Saib, P. Schiavone, T. Figueiro, “Method for Determining the Parameters of an IC Manufacturing Process”, Patent filed, 14306241.2, 2014.

[SAIB, 2014c] M. Saib, A. Fay, P. Schiavone, T. Figueiro, “Method for Calculating the Metrics of an IC Manufacturing Process”, Patent filed, 14305834.5, 2014.

[SALTELLI, 2000] A. Saltelli, K. Chan, E. M. Scott, “Sensitivity Analysis”, Wiley Series in Probability and Statistics”, Wiley, 2000.

[SALTELLI, 2004] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, “Sensitivity Analysis in Practice – A Guide to Assessing Scientific Models”, Wiley, 2004.

[SALTELLI, 2005] A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, “Sensitivity analysis for chemical models”, Chemical reviews, Vol. 105, 2811-2828, 2005.

[SALTELLI, 2008] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, “Global Sensitivity Analysis – The Primer”, Wiley, 2008.

[SCHEFZYK, 2010] D. Schefzyk, D. E. F. Biesinger, D. A. Wharam, “Extraction of the point-spread function in electron-beam lithography using a cross geometry”, Microelectronic Engineering 87, 1091-1094, 2010.

[SCHIAVONE, 2011] P. Schiavone, T. Figueiro, “Procédé de correction des effets de proximité électronique utilisant des fonctions de diffusion décentrées”, Patent filed, 1157338, 2011. (in French)

[SCHLESINGER, 1979] S. Schlesinger, “Terminology for model credibility”. Simulation, Vol. 32, No. 3, pages 103-104, 1979.

[SCHNATTINGER, 2006], T. Schnattinger, E. Bär, A. Erdmann, “Mesoscopic resist processing simulation in optical lithography”, Proc. of IEEE, International conference on simulation of semiconductor processes and devices, 341-344, 2006.

[SCHWARZ, 1978] G. Schwarz, “Estimating the dimension of a model”. Ann. Statistics 6, 461-464, 1978.

[SCZYRBA, 2010] M. Sczyrba, C. Romeo, F. Schurack, T. Castro, B. Connolly, “How to match without copying: an approach for APSM mask process matching using aerial imaging”, Proceeding of SPIE 7823, 2010.

[SEIFZADEH, 2011] S. Seifzadeh, M. Rostami, A. Ghodsi and F. Karray, “Parameter Selection for Smoothing Splines using Stein’s Unbiased Rism Estimator”, Proc. Of. International Joint Conference on Neural Networks (IJCNN), 2733 -2740, 2011.

[SEO, 2006] S. Seo, “A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets”, University of Pittsburgh, Master Thesis, 2006.

[SILVERMAN, 2002] P. J. Silverman, “The Intel Lithography Roadmap”, Intel Technology Journal, Vol. 6, Issue 2, 55-61, 2002.

[SOBOL’, 1967] I. M. Sobol’, “Distribution of points in a cube and approximate evaluation of integrals”. Comput. Maths. Math. Phys. 7, 86–112, 1967.

[SPEARMAN, 1904] C. Spearman, “The proof and measurement of association between two things”. Journal of Psychology 15, 72-10 1, 1904.

[STEVENS, 1986] L. Stevens, R. Joncjheere, E. Froyen, S. Decoutere, D. Lanneer, “Determination of the proximity parameters in electron beam lithography using doughnut-structures”, Microelectronic Engineering, Vol. 5, Issues 1-4, 141-150, 1986.

[STOBERG, 2013] I. Stoberg, D. Dunn, “Etch correction and OPC, a look at the current state and future of etch correction”, Proceedings of SPIE Vol. 8685, 2013.

[STOLBERG, 2010] I. Stolberg, L. Pain, J. Kretz, M. Boettcher, H. Doering, J. Gramss, P. Hahmann, “Electron Beam Direct Write – Shaped Beam Overcomes Resoluton Concerns”, 23rd European Mask and Lithography Conference, Proceedings of SPIE Vol. 6533, 2007.

[SUNDBERG, 2010], L. K. Sundberg, G. M. Wallraff, A. M. Friz, B. Davis, A. E. Zweber, R. Lovchik, E. Delamarche, T. Senna, T. Komizo and W. Hinsberg, “Two Complementary Methods to Characterize Long Range Proximity Effects due to Develop Loading”, Proceedings of SPIE Vol. 7823, 2010.

[SZEKELY, 2007], G. J. Székely, M. L. Rizzo, N. K. Bakirov, “Measuring and testing independence by correlation of distances”, Annals of Statistics, 35/6, 2769-2794, 2007.

[TANABE, 2010] H. Tanabe, T. Abe, Y. Inazuki, N. Hayashi, “Short-range electron backscattering from EUV masks”, Proceedings of SPIE 7748, 2010.

[TARANTOLA, 2006] S. Tarantola, D. Gatelli, T. Mara, “Random balance designs for the estimation of first order sensitivity indices”, Reliability Engineering and System Safety, Vol. 91, 717-727, 2006.

[TEJNIL, 2008] E. Tejnil, Y. Hu, E. Sahouria, S. Schulze, M. J. Tian, E. Guo, “Advanced Mask Process Modeling for 45-nl and 32-nm Nodes”, Proc. of SPIE Vol. 6924, 2008.

[THEIL, 1961] H. Theil, “Economic Forecasts and Policy”. 2nd Edition, North-Holland, Amsterdam, 1961.

[THRUN, 2013] X. Thrun, K-H. Choi, N. Hanisch, C. Hohle, J. Steidel, D. Guerrero, T. Figueiro, J. W. Bartha, “Effects on electron scattering and resist characteristics using assisting underlayer for e-beam direct write lithography”, Proceedings of SPIE 8682, 2013.

[TOLLKÜHN, 2006] B. Tollkühn, M. Uhle, J. Fuhrmann, K. Gärtner, A. Heubner and A. Erdmann, “Benchmark of a lithography simulation tool for next generation applications”, Microelectronic Engineering, 83, 1142-1147, 2006.

[TOP, 2011] M. K. Top, “Analyse des modèles résine pour la correction des effets de proximité en lithographie optique”, Université de Grenoble, PhD Thesis, 2011. *(in French)*

[TORRES, 2003] C. M. S. Torres, S. Zankovych, J. Seekamp, A. P. Kam, C. C. Cedeño, T. Hoffmann, J. Ahopelto, F. Reuther, K. Pfeiffer, G. Bleidiessel, G. Gruetzner, M. V. Maximov, B. Heidan, “Nanoimprint lithography: an alternative nanofabrication approach”, Materials Science and Engineering: C, Vol. 23, Issues 1-2, 23-31, 2003.

[TORTAI, 2012] J-H. Tortai, P. Schiavone, T. Figueiro, N. Jedidi, “Procédé de correction des effets de proximité électronique utilisant des fonctions de diffusion de type voigt”, Patent filed, 1253389, 2012. *(in French)*

[TSAI, 2013] H. Y. Tsai, M/ Guillorn, G. Doerk, J. Cheng, D. Sanders, K. Lai, C-C. Liu, M. Colburn, “Directed self-assembly for ever-smaller printed circuits”, SPIE Newsroom, 2013.

[TU, 2012] S. Tu, L. Xu, “A theoretical investigation of several model selection criteria for dimensionality reduction”, Pattern Recognition Letters, Vol. 33, 1117-1126, 2012.

[TUKEY, 1977] J. W. Tukey, “Exploratory Data Analysis”, Addison-Wesley, 1977.

[ULRICH, 2010] A. Ullrich, “Detecting measurement outliers: remeasure efficiently”, Proceedings of SPIE 7823, 2010.

[UNAL, 2011] N. Unal, M. D. B. Charlton, Y. Wang, U. Waizmann, T. Reindl, U. Hofmann, “Easy to adapt electron beam proximity effect correction parameter calibration based on visual inspection of a “Best Dose Sensor”, Microelectronic Engineering 88, 2158-2162, 2011.

[VILLA-VIALANEIX, 2012] N. Villa-Vialaneix, M. Follador, M. Ratto, A. Leip, “A comparison of eight metamodeling techniques for the simulation of N20 fluxes and N leaching from corn crops”, Environmental Modelling and Software 34, 51-66, 2012.

[VILLARRUBIA, 2005] J. S. Villarrubia, B. D. Bunday, “Unbiased Estimation of Linewidth Roughness”, Proceedings of SPIE 5752, 480-488, 2005.

[WELLS, 1999] R. J. Wells, “Rapid approximation to the Voigt/Faddeeva function and its derivatives”, Journal of Quantitative Spectroscopy and Radiative Transfer, Vol 62, 29-48, 1999.

[YAMAMOTO, 2000] J. Yamamoto, F. Mural, Y. Someda and S. Uchino, “Fine Pattern Fabrication below 100nm with 70kV Cell Projection Electron Beam Lithography”, Jpn. J. Appl. Phys., 39, 6854–6860, 2000.

[YAMAZAKI, 2000] K. Yamazaki, M. Z. M. Zaifullah, H. Namatsu and K. Kurihara, “Sub-10-nm electron beam lithography with sub-10-nm overlay accuracy”, Proceedings of SPIE 3997, Emerging Lithographic Technologies IV, 458, 2000.

[YANG, 2005] Y. Yang, “Can the Strengths of AIC and BIC Be Shared? A conflict between model identification and regression estimation”, Biometrika, Vol. 92, No. 4, 937-950, 2005.

[YEH, 2006] K. Yeh, C. Lin, J. Hu and W. Loong, “Simulation for Optimization of Mask Error Enhancement Factor by Design of Experiments in Both Dry and Immersion ArF Lithography”, Jpn. J. Appl. Phys., 45, 6216–6224, 2006.

[YE, 1998] K. O. Ye, “Orthogonal column Latin hypercubes and their application in computer experiments” Journal of the American Statistical Association, vol. 93, 1430-1439, 1998.

[YOSHIZAWA, 2004] M. Yoshizawa, K. Iwase, H. Ohtorii, K. Oguni, H. Hane, K. Amai, S. Moriya, H. Nakano and T. Kitagawa, “Analysis of critical dimension uniformity for LEEPL”, Emerging Lithographic Technologies VIII, Proc. Of SPIE Vol. 5374, 2004.

[ZERBE, 2007] E-M. Zerbe, T. Marschner, J. Richter, C. Utzny, “Matching of Different CD-Metrology Tools for Global CD Signature on Photomasks”, Proceedings of SPIE Vol. 6607, 2007.

ANNEXES

“Everything should be made as simple as possible, but not simpler”

- *Albert Einstein*

Annex I - Lithography Characteristics Metrics

There are some parameters that are useful to quantify the characteristics of the exposure techniques for writing patterns. These parameters are useful for both evaluating the limits of a given process and for comparing different processes. The most common concepts are presented in the following subsections.

Critical Dimension (CD)

The Critical Dimension (CD) corresponds the smaller dimension of a given pattern. In many contexts the CD definition may be extended to represent any dimension worth measuring.

Dose to Size (DTS)

For most of the lithography techniques explained in this chapter, each pattern is exposed by a given dose (the notable exception is the nanoimprint). Even if it is measured in mJ/cm² (optical lithography) or in µC/cm² (e-beam lithography), there is a given amount of energy that will be required to print the pattern in their target dimension. The dose required to print a given pattern at the desired dimension is called Dose to Size (DTS).

Energy Latitude (EL)

In the specification of a process it is usually included a tolerance on variation among CD values. This tolerance, expressed either in an absolute value in nm or in a relative value to the "pattern CD". Therefore, for a given pattern, the Energy Latitude (EL) is the tolerance expressed in terms of dose variation. It means that, for a given CD the EL indicates the maximum energy fluctuation around the DTS that will keep the dimension of the pattern inside the tolerance range [YAMAMOTO, 2000].

Depth of Focus (DoF)

Depth of Focus (DoF) is the tolerance of placement error on the image plane in the z direction relative to the lens. An approximate expression of the depth of focus is given by (3). In optical lithography it is one of the major factor to limit the resolution of the lithography process. On the other hand, DoF for e-beam machines is much relaxed and is not considered as a limiting factor. In e-beam it is common to replace the notion of DoF by beam to beam dispersion

Mask Error Enhancement Factor (MEEF)

The Mask Error Enhancement Factor (MEEF) allows the characterization of the sensitivity of a given process to the errors presented on the mask [YEH, 2006], [YOSHIZAWA, 2004]. For a given CD, the way to determine the MEEF is to vary the dimension on the mask and observe the corresponding variation on the wafer. MEEF is computed as shown in (96)(96).

$$MEEF = \frac{\Delta CD_{wafer}}{\Delta CD_{mask}} \tag{96}$$

Although the MEEF is impacted by several characteristics of the lithography process, it is also dependent of the pattern being analyzed.

Writing Error Enhancement Factor (WEEF)

As the MEEF indicates the impact of mask errors over the final result on the wafer, the Writing Error Enhancement Factor (WEEF) accounts for a similar effect on maskless e-beam lithography [LIN, 2010]. The reference element for WEEF is the beam fluctuation. The way to verify this impact is to slightly vary the pattern dimension and observe its impact on the wafer.WEEF is computed as shown in (97).

$$WEEF = \frac{\Delta CD_{wafer}}{\Delta CD_{layout}} \tag{97}$$

CD Uniformity (CDU)

In order to quantify the dispersion of CD values over a mask or a wafer, it is usual to compute the CD Uniformity (CDU) [POSTNIKOV, 2003]. The most frequent way to evaluate CDU is to measure the same pattern several times in different locations and then compute the 3σ value from the measurements. This is a metric of the robustness of the lithography process and has a direct impact on the yield of the process.

Line Width Roughness (LWR) and Line Edge Roughness (LER)

The patterns as exposed on the wafer usually present some variation on its edges. This variation, of statistical nature, is usually quantified as the uniformity of the "pattern CD" for a given length and is called Line Width Roughness (LWR) or as the uniformity of a single edge of the pattern, also for a given length, and is called Line Edge Roughness (LER). Figure 141 shows a representation of the definitions of LWR and LER [VILLARRUBIA, 2005].

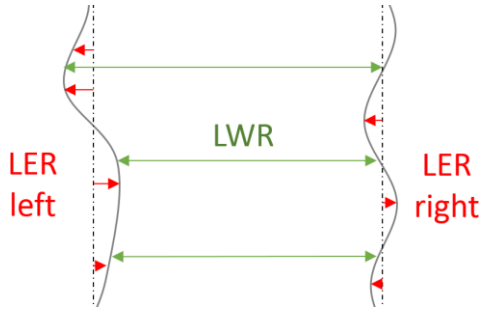


Figure 141. Representation of the definitions of LER (in red) and LWR (in green).

Usually the value is represented in terms of 3σ, where σ is the standard deviation of the distances between the two edges (for LWR) or between the edge and the mean value (LER).

This evaluation is performed both on the wafer level and on the mask level, for different reasons. At wafer level, controlling the roughness means improving circuit performance since these variations impact the electronic behavior of the features, such as transistors. On the mask level, LER is considered to be a significant contributor to the overall feature CDU, especially for small features.

Line End Shortning (LES)

It is common while printing lines on a lithography process that they may print shorter than intended in the pattern design. This phenomenon is known as Line End Shortening (LES). It is

an important issue because as the layers of a circuit must connect, the overlaying area is impacted by the shorting reducing or removing completely the intersection of the layers.

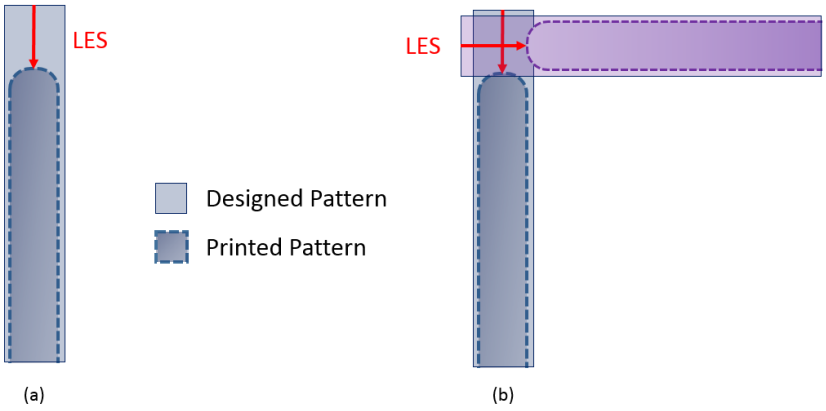


Figure 142. (a) Differences between the designed pattern and the printed one due to LES and (b) a representation of what kind of issue a design may face due to LES.

Annex II - Exposure Compensation

From the early stages of e-beam lithography used for writing masks or wafers, the lack of precision in the e-beam exposure were orders of magnitude below the feature sizes. As technology evolved and the feature sizes got smaller, the industry adopted a set of rules to improve the exposure results in a first order, which was sufficient at the time. This approach is known as rule-based correction (or compensation) and may exemplified as a set of rules such as:

- For CDs smaller than 500nm, increase the exposure dose by 10%;
- For regions with patterns covering more than 50%, decrease the exposure dose by 20%;
- Etc.

These rules were getting more complex at every new technology node, as well as more exceptions and precisions were required, up to the point that the compensation was non practical without the use of a model to predict how the exposure result would be.

As described in the first chapter of this thesis, e-beam resolution is mitigated by various phenomena such as electron scattering, acid diffusion in the resist, etching effect, fogging effect, among others. In this section some of the most frequently used strategies for compensating these phenomena are presented.

Dose Modulation

In a general way, proximity effect correction on e-beam exposure tools intend to compensate for the difference between the energy targeted to the material and the one received in the end. This difference comes from the forward scattering phenomena and the backscattering phenomena. Other phenomena, such as resist diffusion, etching and others are also compensated, but in an indirect way.

The most straightforward way to compensate for this energy difference is to change the amount of energy sent to each pattern. This strategy is called dose modulation and is better illustrated by following examples.

Consider the exposure of an isolated pattern. The energy sent to expose the pattern scatters forward to regions that were not supposed to receive energy. Though the energy may not be enough to sensitize the resist in these areas, it impacts the resolution of the pattern by lowering the real dose received where the pattern is supposed to be printed. Figure 143 (a) illustrates the example where no modulation is performed. If a threshold of the resist (TH) is set at a given dose, one can see that the targeted CD is not achieved. Figure 143 (b) shows the dose modulation example (here an increase of the total dose is performed) where the real dose is the targeted one at least in what concerns the resist threshold and, therefore, the targeted CD is obtained.

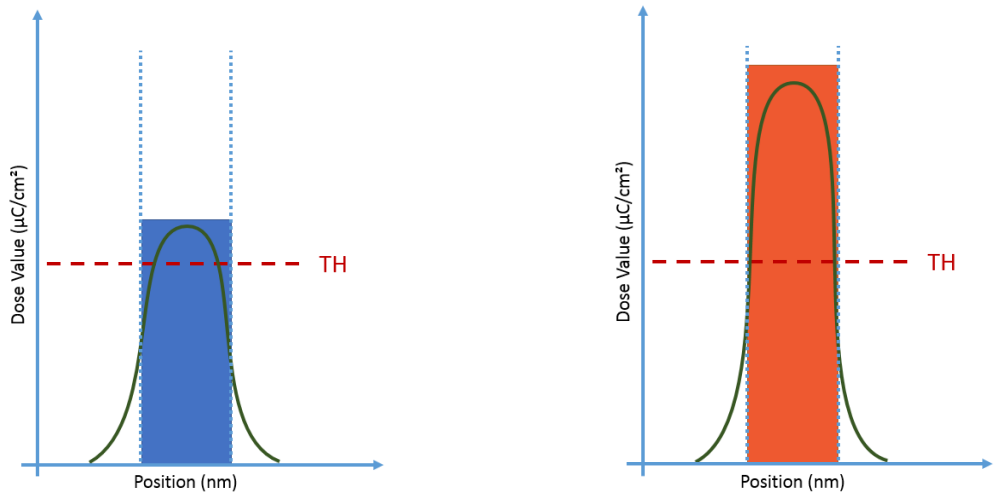
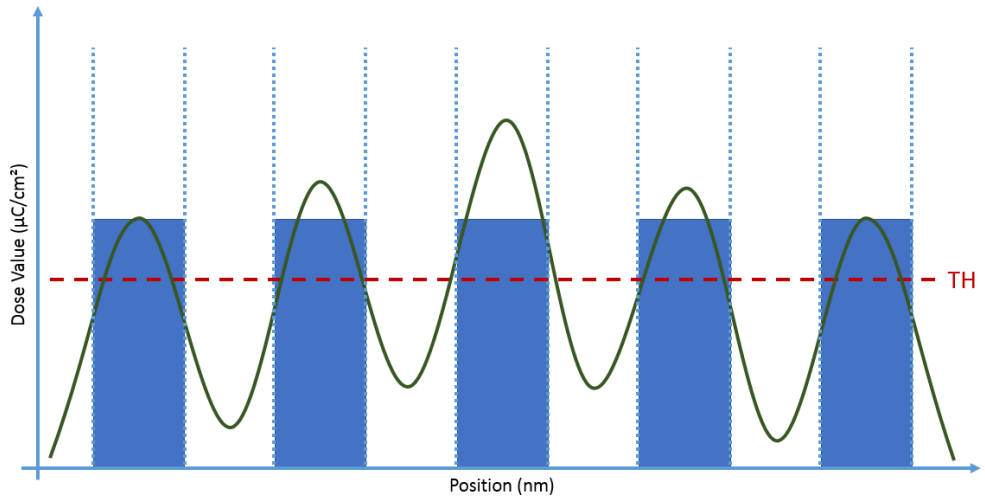
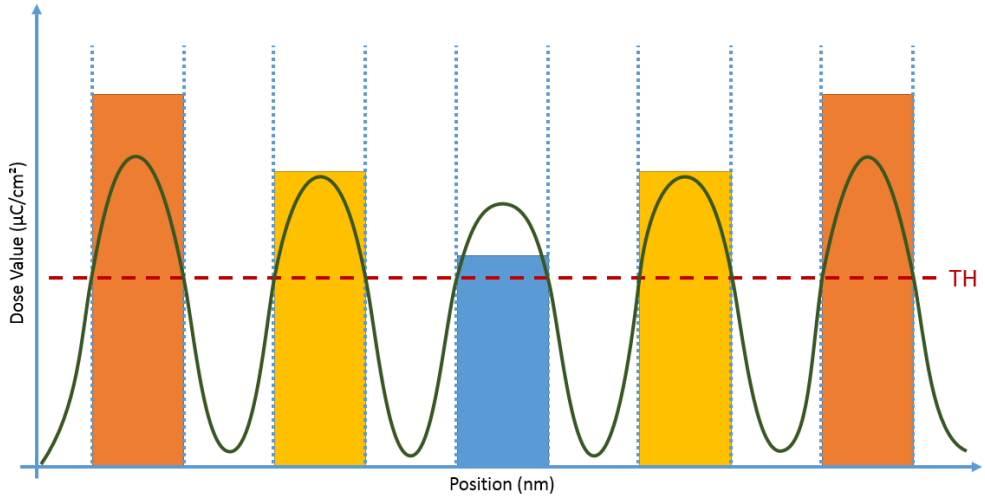


Figure 143 Example of a side view of an isolated line with (a) a standard dose and the CD not met at the threshold (TH) level of the resist and (b) after dose modulation, achieving the CD value at the TH level.

In a second example, consider the exposure of line/space (L/S) features with the pitch equal to two times the CD. The energy from backscattered electrons coming from the exposure of the other patterns increases the dose in the center pattern. Side patterns do not receive as much energy as patterns in the center. In this case, there is no single dose that can set all patterns to targeted CD. Hence, dose modulation is performed pattern by pattern to match the resist threshold (TH) with the proper CD. Figure 144(a) illustrates this example of a L/S features block using a single standard dose. Pattern dimensions depend on pattern position and Figure 144(b) presents the dose modulation that allows reaching the targeted CD for each pattern.



(a)



(b)

Figure 144 Example of a side view of L/S features with (a) a standard dose and the CDs of the patterns not met at the threshold (TH) level of the resist and (b) after dose modulation, achieving the CD value at the TH level for each pattern by using different doses.

Multiple Exposures

Instead of performing a dose modulation over the shapes to be exposed one may consider performing multiple exposures. This strategy consists in determining an initial dose for the original shapes and then adding features that increase locally the dose received by the original pattern in order to improve lithography resolution. This strategy not only allows obtaining the right CD but also gives more control in order to improve other characteristics such as LER and EL [Martin, 2009]. Figure 145 shows an example of how combining the assist features to the original one improves energy latitude and allows reaching the target CD at the TH level of the resist.

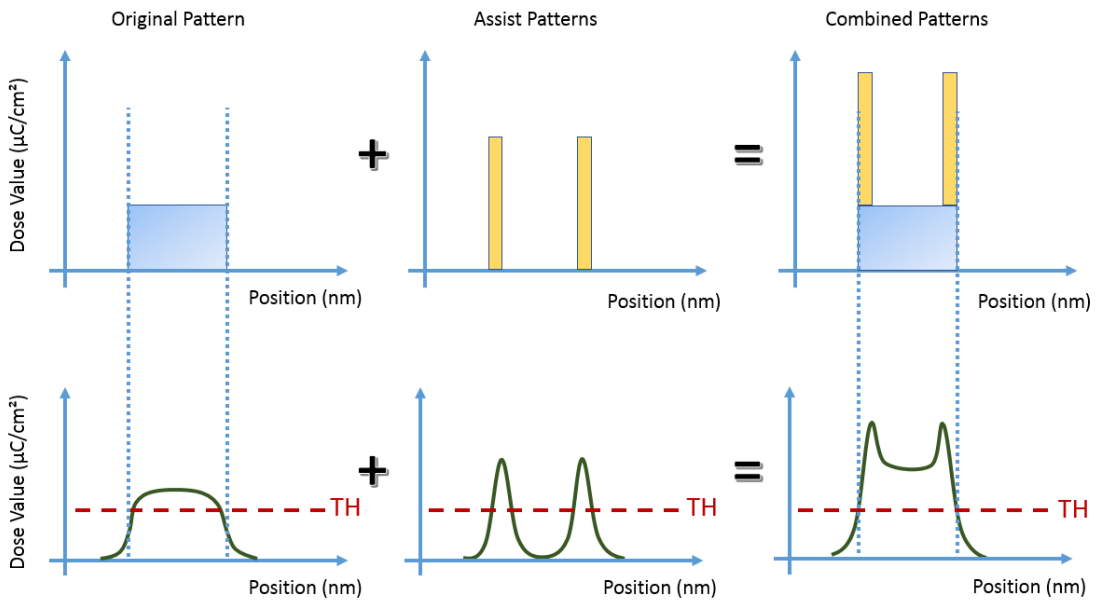


Figure 145 Example of adding assist features to obtain the target CD.

Although this method may accurately reach the CD targets, it increases the exposure time by adding more features to the layout. However, this strategy also increases the LER and reduces the EL due to its high background dose levels.

Geometry Modulation

Another strategy to compensate for the difference between the energy sent to the material and the one received in the end is to change the exposed region. An increase in the size of the exposed region increases the total dose sent to the pattern and, if properly set, may allow the right amount of energy to be exact to meet the pattern intended dimension.

Since this energy difference is translated into a dimension difference from the target CDs, one may consider changing the exposed patterns sizes in such a way that the final result meets the original target CDs. This strategy is called geometry modulation and is better illustrated by the following examples.

Consider the same example of exposure of an isolated pattern, as presented in 8. The energy received in the region where the pattern is supposed to print is not enough. Instead of increasing the dose value sent to the pattern, one may increase the pattern size in the sense that, after the "reduction" coming from the forward scattering, it actually meets the target CD. Figure 146(a) illustrates the example where no geometry modulation is performed and the CD is not achieved in the threshold of the resist (TH) while Figure 146 (b) shows the geometry modulation example where the CD is met at the threshold level.

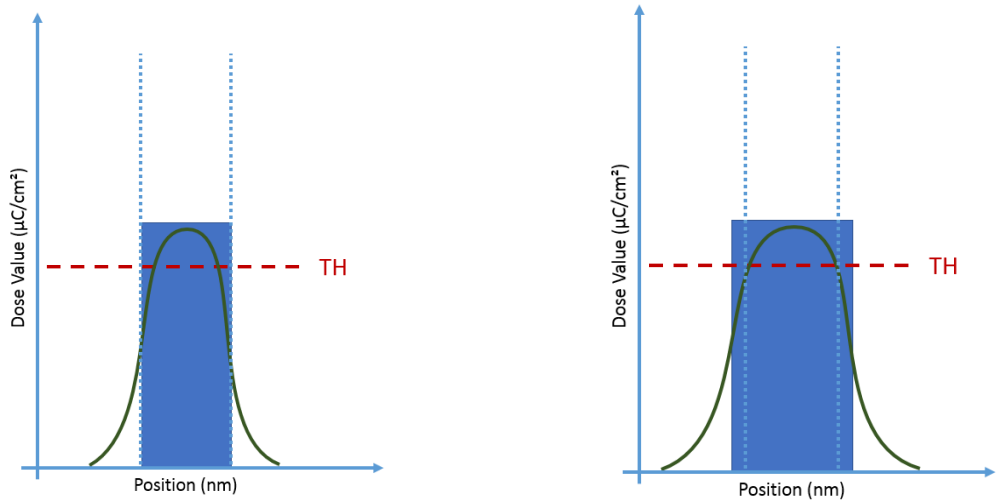


Figure 146 Example of a side view of an isolated line with (a) its original dimension and the CD not met at the threshold (TH) level of the resist and (b) after geometry modulation, achieving the CD value at the TH level.

In a second example, consider the exposure of line/space (L/S) features identical as the one presented in 8. A geometry modulation strategy may be applied in order to fix all patterns by increasing the ones at the edge in order to increase the overall received dose and decreasing the dose in the patterns located in the center. Figure 147 illustrates this example of a L/S features block (a) using a single standard dose that is not reaching the target CD while (b) presents the dose modulation result, reaching the CD for each pattern.

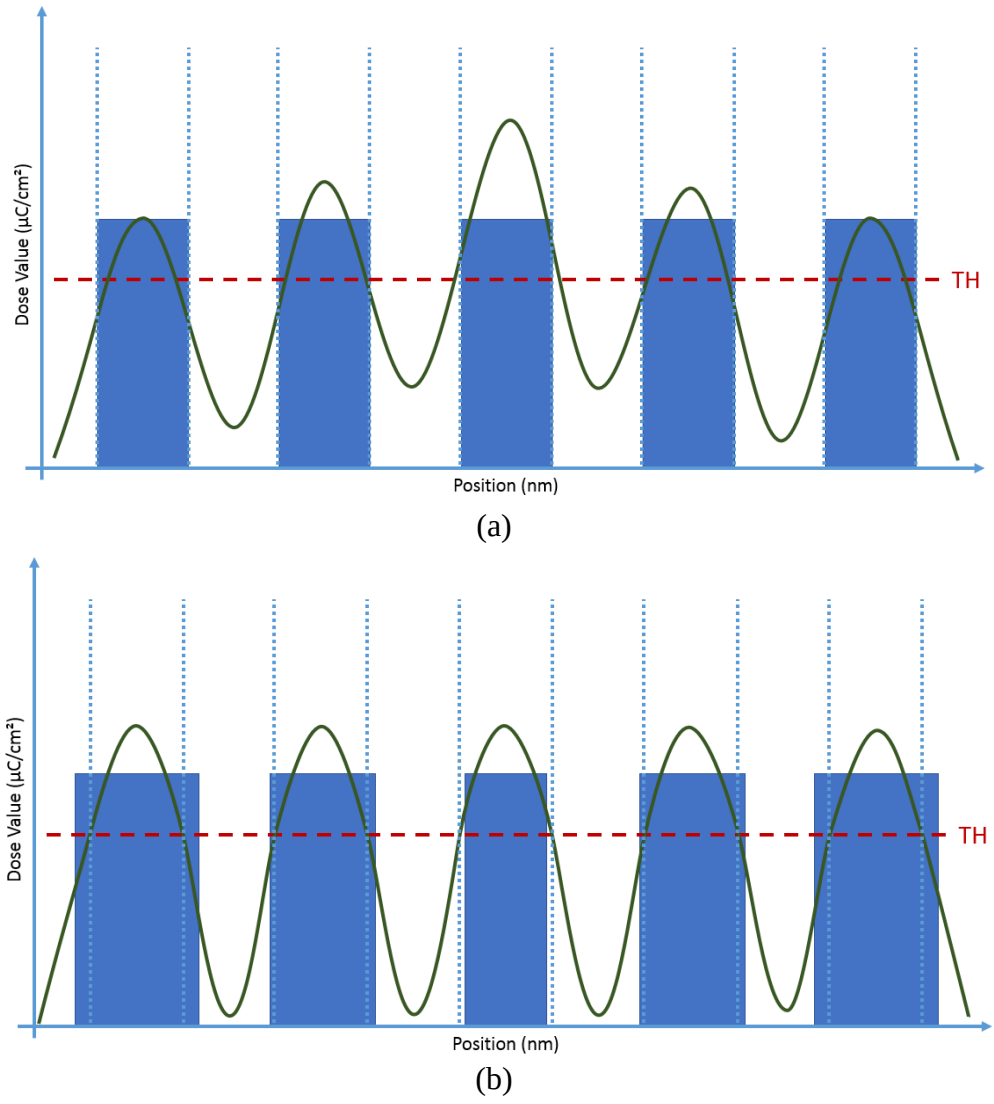


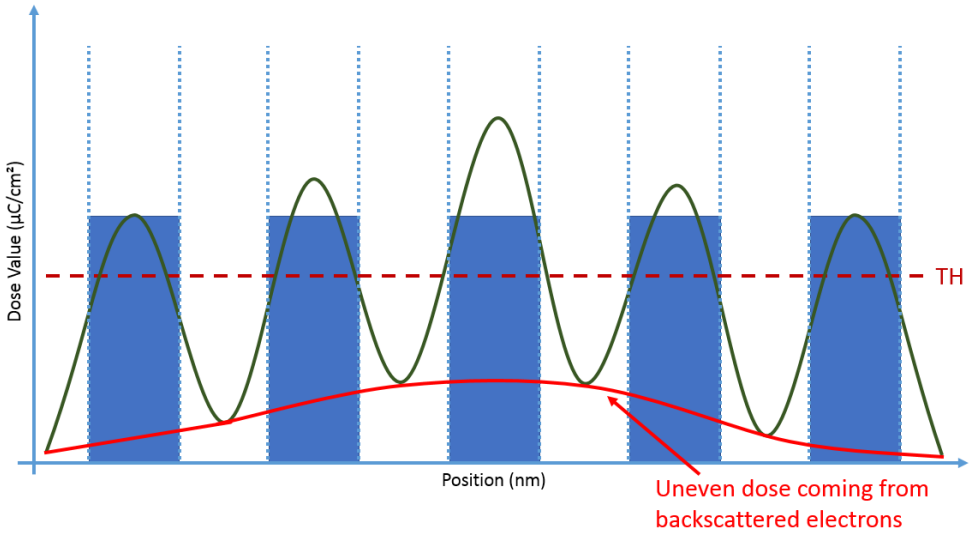
Figure 147 Example of a side view of L/S features with (a) theirs original dimension and the CDs of the patterns not met at the threshold (TH) level of the resist and (b) after geometry modulation, achieving the CD value at the TH level for each pattern by using different doses.

Ghost Method

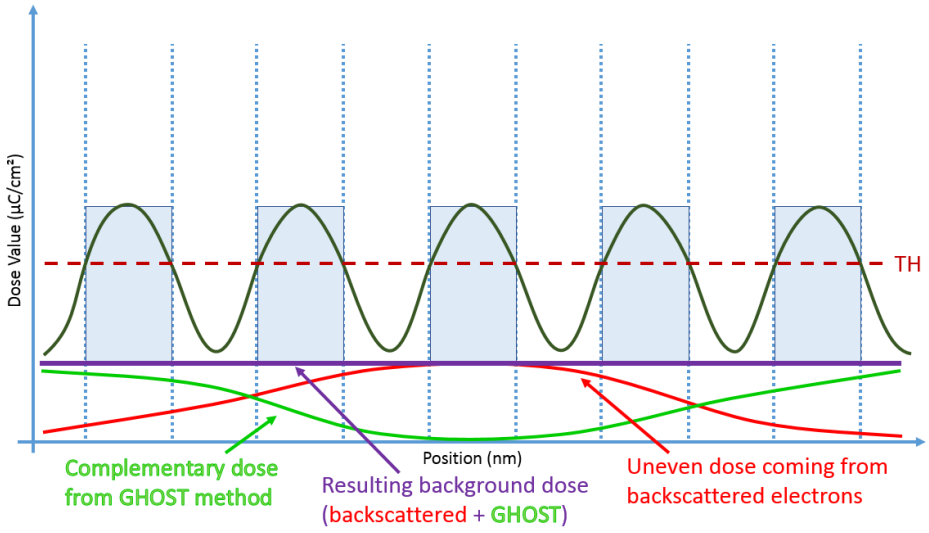
Another strategy is known as the "GHOST" method which is based in performing a second exposure of all regions where the density is too low. In this sense, the strategy regards the energy from backscattering phenomena as a background dose that exist everywhere but in different intensities. The way to compensate for it is to even the dose for the entire chip by performing an exposure over the regions with lower densities.

This process works in two exposures, the first being responsible for writing the patterns with the estimated dose they require considering a uniform background dose and a second exposure that is actually the complement of the estimated dose coming from backscattered electrons. The second exposure is usually performed in a pixelated background that receives the complementary energy for the specific region. Figure 148 shows the example of applying the GHOST method over the L/S features from previous examples. Observe in Figure 148 (a) the red line identifying the energy considered to come from backscattered electrons or any other phenomena related to the density and in Figure 148 (b) the green line that shows the dose

complement required to present a uniform background dose (presented in purple). Observe as well that the dose sent to each pattern is considerably lower than before, to account to the background dose as if all patterns were in high density regions.



(a)



(b)

Figure 148 Example of a side view of L/S features with (a) theirs original dimension and doses, indicating the energy that may be considered coming from the backscattered electrons and (b) after applying the GHOST method to uniform the background dose, achieving the CD value at the TH level for each pattern.

Although this method may accurately reach the CD targets, it presents several drawbacks. The first one is to require two exposures in a method that struggles to improve its throughput. Moreover, this strategy also increases the LER and reduces the EL due to its high background dose levels.

Dose and Geometry Modulation

In recent years the industry came with another strategy to compensate for the e-beam exposure issues that performs, at the same time, dose and geometry modulations [MANAKLI, 2007]. The main advantage of this strategy is to combine the strength of these two methods allowing to compromise between their weak points. For instance, geometry modulation is limited by the

space available for performing its dimensions change while dose modulation may be limited by the highest dose used for not increasing the writing time, for instance.

The approach of combining dose and geometry is currently the state-of-the-art of proximity effect correction for advanced nodes. Besides given flexibility and extending the capabilities of each of the previous approaches, it also gives control for improving other characteristics, such as EL.

Summary

The different strategies discussed in this chapter allows exposing patterns with the required changes to meet the desired dimensions. Nonetheless, if the prediction is not accurate enough, the compensation based only on this prediction won't be able to meet the dimensions. Even worse, a wrong prediction may increase the error, by compensating for something that was happening in the opposite way.

Annex III - Calibration Design Floorplanning

The way to distribute the patterns over the calibration layout may impact the calibration procedure itself. The phenomena under modeling, and specially their range, must be taken into consideration while deciding the dimensions of the patterns and their placement on the layout.

In terms of the pattern dimensions, consider the following example. If an L/S block is supposed to provide a uniform density around its center, one should consider creating a pattern large enough to ensure that the center point has a uniform surrounding. Figure 149 illustrates the minimum dimensions of a calibration pattern in order to ensure homogeneity of the effects over a given metrology point.

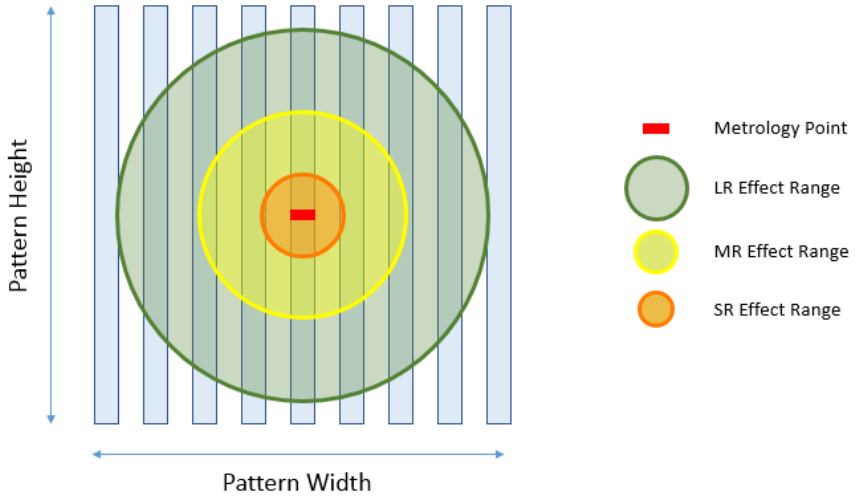


Figure 149 Example of pattern dimensions according to the effects under consideration.

Moreover, if measurements are supposed to be performed off-centered of the pattern, one should also consider spacing different calibration patterns inside the calibration design in order to prevent them to interact. Figure 150 illustrates the minimum distance among different calibration patterns to prevent interaction among them and the metrology points.

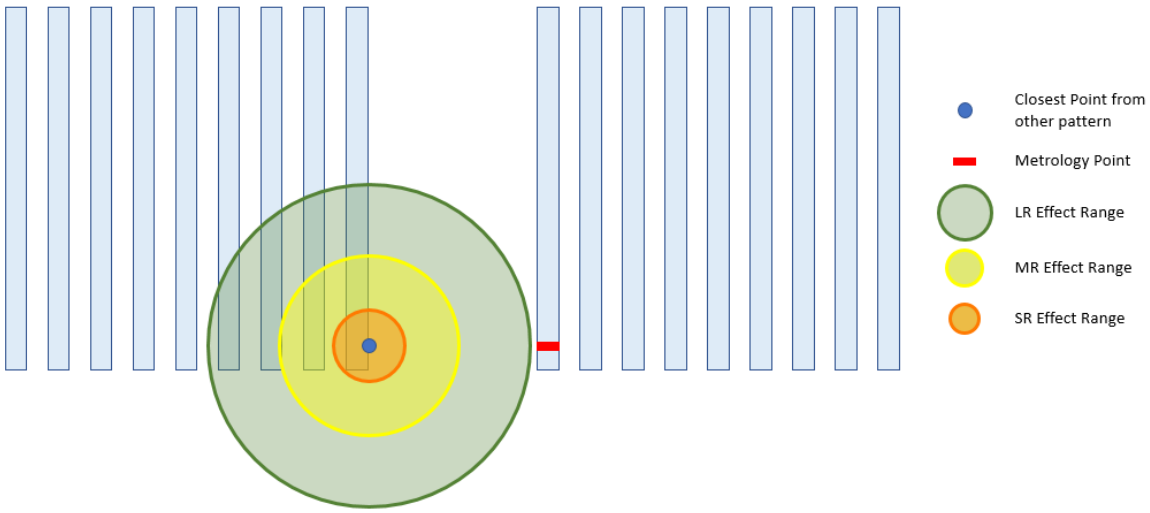


Figure 150 Example of distance between patterns according to the effects under consideration.

Finally, for metrology purposes, it is usually more convenient to place the metrology points into a matrix of regular pitch in both directions. This allows the creation of the metrology recipes to be simpler and more precise. For this reason, most of the standard calibration designs look like the illustration presented in Figure 151.

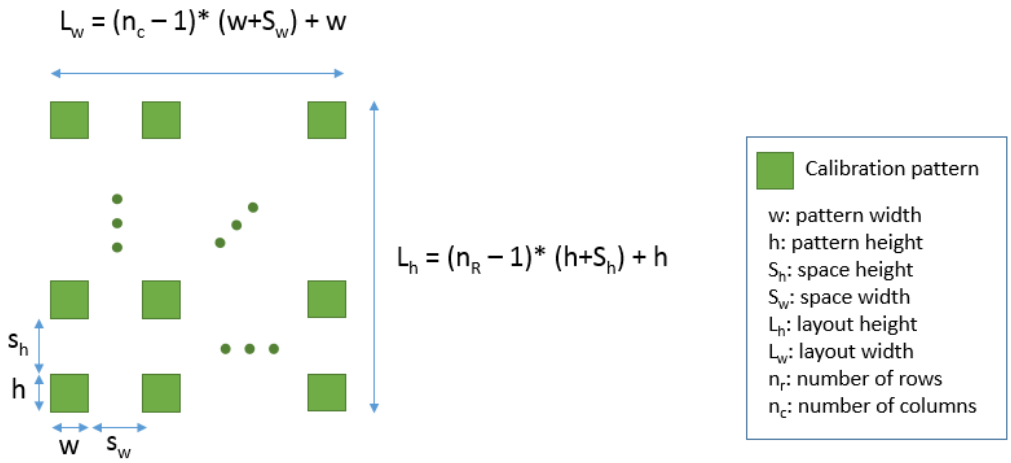


Figure 151 Illustration of a calibration design floorplanning.